CHAPTER ONE

Introduction: Transferring Ideas from One Brain to Another

1.1 The Speech Chain

If you read a story about a superhero who could transmit their thoughts into another person's brain without any physical contact, you might think it was a little far-fetched. But in fact, that's exactly what happens during spoken communication: I can take an idea from my brain and place it in your brain. Of course, this brain-to-brain transfer does not require mystical powers – it happens using our body's ability to produce and detect subtle variations in air pressure, and our brain's exquisite ability to harness these processes. But the overall feat remains remarkable, even though we experience it on a daily basis.

Language itself is fundamental to human experience, both as individuals and as a species. But language is also intimately related to other domains of science that have historically been studied in isolation, including memory, emotion, sensory processing, decision-making, and motivation (to name a few). The neuroscience of language is thus a useful microcosm of human experience that transcends traditional boundaries between disciplines. As we will return to in Chapter 10, language is truly a whole-brain activity.

Although I touch on other forms of language processing, the focus of this book is on spoken verbal language – that is, on human speech. Various forms of vocal communication are present in many species, and in humans spoken communication is evolutionarily older than written communication. Many theoretical considerations are similar across different modalities of language, including spoken, written, and signed language. Thus, a firm grounding in spoken communication provides a vocabulary of terms and ideas that translate well to other aspects of language. However, it is important to recognize there are many different modalities of human language, of which speech is just one (a theme we return to in Chapter 9).[1]

A cartoon overview of spoken communication, modeled after the original "speech chain" from Denes and Pinson (1993), is shown in Figure 1.1, and loosely informs how this book is organized. Spoken communication begins with an idea in a talker's head, which is then translated into acoustic vibrations using their mouth and vocal folds (a process that involves monitoring their own production via auditory and somatosensory input). These sound waves travel

Cambridge University Press & Assessment
978-1-009-24527-2 — The Neuroscience of Language
Jonathan E. Peelle
Excerpt
More Information

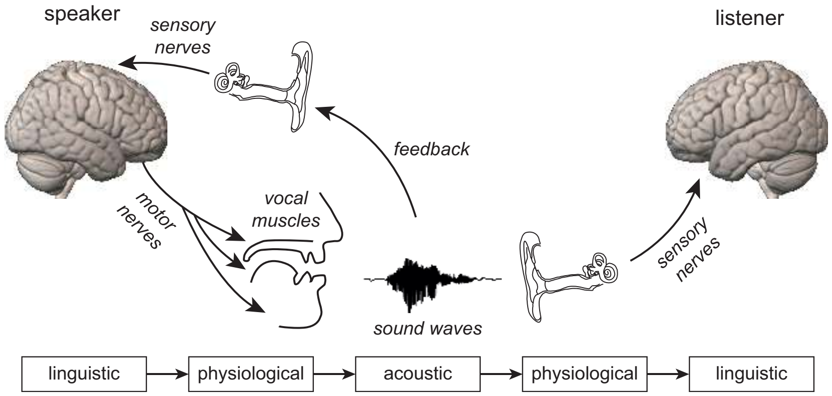

Figure 1.1 Schematic of the "speech chain" (after Denes and Pinson, 1993). To transfer an idea to a listener, a speaker controls their body in a specific way to produce air vibrations that we experience as sound. These travel to the auditory system of both the listener (who will, hopefully, extract the intended meaning) and the speaker (who can check to make sure they produced the sounds they intended). https://osf.io/geqb6/ (CC-BY).

through the air and may be mixed with environmental noise or otherwise degraded before reaching the ear, and auditory system, of the listener. After traveling up a complex and interconnected ascending auditory pathway, the now-transformed acoustic signal is processed by cortical auditory and speech regions to extract meaning (which we experience as "understanding" what a talker has said). Given the rapid rate of speech, the above process must happen very quickly, and also continuously (at least during extended listening, such as in a conversation or lecture). Throughout the rest of the book, we will examine these stages of the speech chain in greater detail.

1.2 Levels of Speech Processing

Linguistics is the study of language; branches of linguistics deal with topics including speech sounds, concepts, how words are combined, and comparisons of language attributes across cultures. Psycholinguistics specifically deals with how the mind processes language – the psychological facets of linguistics – and thus focuses on the perceptual, cognitive, and linguistic operations required for language. Researchers interested in the neuroscience of language are influenced by both of these traditions, as well as by thinking in neuroscience and psychology more broadly. To understand the terminology in

Table 1.1 Key levels of speech processing.

phoneme	Smallest meaningful speech unit.
syllable	Unit of speech that combines to form words (for example, a consonant–vowel pair).
morpheme	A meaningful linguistic unit that can't be divided up into smaller units. A word like "dog" is a morpheme; so is the "s" at the end of "dogs." A free morpheme can occur as a separate word, whereas a bound morpheme cannot stand on its own.
lexical	At the level of a word, including word meanings and grammatical function.
sentence	A grammatical unit composed of one or more clauses.
discourse	Language unit longer than a single sentence, including the function of language in conversation more broadly.

the rest of the book and research papers you read, a preliminary understanding of some basic terms will be useful. These are listed in Table 1.1.

An enduring challenge to our understanding of how the brain understands language is creating a mapping between terms used in different fields. Sometimes different terms refer to a similar underlying concept; other times *similar* terms refer to *different* concepts.[2] Clear and explicit definitions are useful to keep close at hand.

It is also worth considering how existing linguistic and psycholinguistic frameworks have influenced areas of research in the neuroscience of language. For example, from a language-centered background, it makes sense to identify neural mechanisms underlying phonemes, syllables, and morphemes; from a strict auditory neuroscience perspective, however, it might make more sense to think about amplitude fluctuations and the acoustic complexity of the signal. Combining these perspectives is one of the most challenging facets of the neuroscience language but can also be one of the most rewarding.

1.3 Challenges of Spoken Language Processing

Before we get too far into the neuroscience of language, it's useful to think about some of the challenges listeners need to overcome for successful communication to occur (many, though not all, of which are shared to some degree across different modalities of language and, indeed, across other cognitive domains).

1.3.1 Categorical Perception

Categorical perception refers to the general phenomenon in which things that are technically not identical are perceived as being linked together as the same *kind* of thing. For example, at some point you almost certainly learned what a chair is. In fact, you developed a concept of "chairness" that goes beyond any particular example of a chair that you encountered. As a result, you probably find it easy to recognize whether a piece of furniture is a chair or not, even if you have never seen it before. Importantly, your concept of CHAIR (I use all capitals to denote concepts, rather than particular words) encompasses different types of chair: an office chair, a dining-room-table chair, an easy chair, and so on. You have learned a category of objects that share certain features and label these items "chair" – you don't think, well, this one is 100 percent of a chair but this other one is a little different so I'll say it's just 70 percent chair.

Now, for a speech example. Let's say you have four friends and you ask each of them to say the word "dog." Each of your friends has a different body – specifically, a different vocal tract, vocal folds, mouth, and so on – and thus the acoustic vibrations coming out of their mouth when they say "dog" are going to be different (that is, the acoustic signals are not identical). Nevertheless, as a listener, you will have no trouble understanding that they are saying "dog." In other words, you have mapped four different sounds onto the same category (in this case, "dog"). This is sometimes expressed in terms of a "many-to-one" mapping, because a large number of acoustic sounds need to be associated with a single category (in this case, a word).

The field of speech research has many examples of categorical perception as studied in a laboratory setting. In one classic paradigm, listeners are played a series of sounds and given two choices: For example, did you hear a "pa" or a "ba"? Unbeknownst to the listeners, the sounds they hear vary continuously between "ba" and "pa" (that is, the acoustic signal varies in equal steps between a continuum that moves from 100 percent "ba" to 100 percent "pa"). However, the responses that people make show a classic sigmoidal function (see Figure 1.2), indicating that "ambiguous" sounds (acoustically between "ga" and "ba") are more likely to be heard as *either* "pa" or "ba" – that is, they are perceived as belonging to a category (in this case, listeners of English know that "pa" and "ba" are both possible options). Perhaps most importantly, people struggle to discriminate tokens that fall within a category but *are* able to discriminate tokens that fall between categories. The learning of these speech categories during childhood, and how listeners transform a continuous acoustic space into a discrete word space, have long intrigued speech scientists.[3]

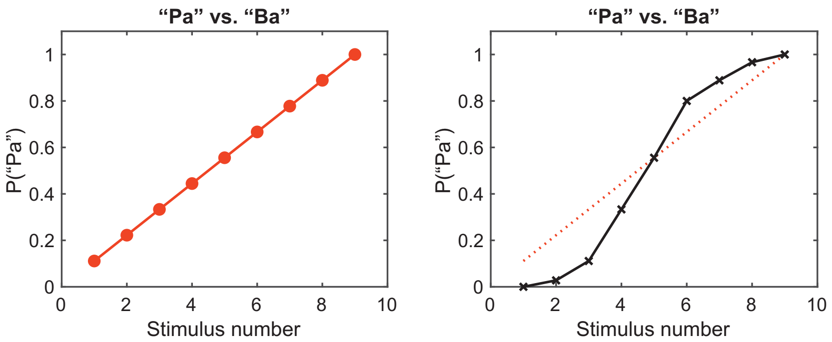

Figure 1.2 Categorical perception of speech is demonstrated by a lack of linearity in how listeners perceive speech. For example, in this cartoon example, the acoustic signal changes in equal steps from "pa" to "ba" (stimulus 1 = 100 percent "ba," stimulus 9 = 100 percent "pa"). The probability of a listener reporting "pa" – P("pa") – is plotted on the y axis. *Left:* If our perception were true to reality, the probability of choosing "pa" or "ba" would also change equally (red line). *Right:* In real life, listeners tend to choose *either* "pa" or "ba," which is why the curve is steeper in the middle than on the ends. Compare the black line (what listeners tend to do) with the dotted red line (equal change in perception). https://osf.io/geqb6/ (CC-BY).

1.3.2 Time-Constrained Understanding

Although estimates vary, spoken conversation generally happens at a rate of around 140–180 words per minute (which corresponds to 2–3 words every second). Speaking rate can be substantially faster than this if a talker has rehearsed, is reading from a prepared text, or the speech has been edited (for example, for a TV show or podcast).[4] Listeners must therefore rapidly map the continuous acoustic signal presented to their auditory system to the words they are interested in understanding. Importantly, in live contexts, a listener does not have the opportunity to slow down the incoming speech: The speaking rate is controlled by the talker, not the listener. The next time you attend any large event with a single speaker (such as a lecture, ceremony, religious service, or similar event), take a moment to consider how quickly the speaker is producing speech, and how in control of this you feel, as a listener. You can do the same thing next time you watch a show.[5] Now compare that experience to what you are doing right now when you read this text. Here, you are controlling your own eye movements across the page. If you don't understand a sentence, you can go back and read it again. For example, if I use an unusual word that makes you lour,[6] you might pause, even if the word is entirely cromulent.[7] Or, if you need a second to pause and think (or read an endnote), you can take it whenever you like. Spoken language is thus fundamentally constrained in time in a way that reading is not.[8]

What are the consequences of these time constraints? The need for rapid analysis pressures listeners to make use of all available cues to aid perception. These might include predictions about upcoming words based on prior content and defaulting to common interpretations when ambiguity arises. The heavy reliance on context and prediction is often helpful but can also hinder understanding when unusual (or at least, unpredicted) utterances are encountered.

1.3.3 Flexibility

In the midst of the drive to rapidly map acoustic signals onto previously learned sound categories comes another challenge: namely, the need to be flexible.

Recall that a challenge of categorical perception is the many-to-one problem that requires a listener to perceive different acoustic signals as belonging to the same speech category. However, *which specific* speech sounds get mapped to a speech category is not as straightforward as one might think. Earlier, we encountered the example of different people saying the same word; already, as a listener, we need to accommodate that variability so that we aren't confused when someone new starts talking (fortunately, you could meet a stranger and probably understand them if they said "dog," even though you had never heard them speak before). Our ability not only to accommodate variability in a general sense but also to adjust our perception to match a specific talker suggests that perceptual categories are not set in stone but have some degree of flexibility. Exactly how listeners implement this flexibility is not always clear.

Furthermore, no matter how flexible we are as listeners, there are some situations in which our understanding of speech will start to falter. For example, consider listening in the midst of background noise (think about a crowded restaurant, sporting event, or having the bad luck of sitting next to an air-conditioning unit during a lecture; listening to someone with an accent that is unfamiliar to you; or listening to digitally altered speech, such as a podcast at 1.5 times normal speed). In such situations you might find yourself missing words and not entirely understanding what a talker is saying. However, in many (though not all) of these cases listeners are able to adjust to challenging speech signals over time, a process usually framed as **perceptual learning** or **adaptation**. There is no clear boundary between "normal" flexibility in listening and perceptual learning, but generally the ease and success of perception plays a role. So, for example, when you understand all of your friends saying "dog," there is no additional adjustment required: Your perceptual system is already accommodating this level of variability and you are very accurate at understanding what you hear. However, if you listen to a podcast at twice normal speed, you might find that after a minute of listening you are able to catch more words than when you started – or you might understand the same amount but find it less effortful – reflecting some perceptual learning.[9]

The overall point is that listeners need to be flexible, both in the moment (to accommodate little variations in speech sounds) and over a longer time period (to accommodate greater challenges to perception).

1.3.4 Multimodal Integration

As noted in Section 1.1, most of this book deals with auditory speech (or more specifically, "auditory-only" speech). As listeners we frequently want to understand language that is solely in the auditory modality (radio, podcasts, phone calls, a passenger in the back seat of our car, to name a few real-life examples). However, in many instances even "auditory" speech occurs across modalities, most commonly involving visual speech information in addition to auditory information. It has long been appreciated that being able to see a talker's mouth can improve speech recognition, especially in the presence of background noise (covered in more depth in Chapter 9). Interestingly, some of the information provided by auditory and visual speech signals is redundant, and some is complementary. At a basic level, auditory and visual information are necessarily separated, because they are processed by different sensory organs, and thus enter the brain through different pathways. There is no question that auditory and visual speech information *are* combined – simple behavioral experiments reliably demonstrate this fact – but the *way* in which they are combined continues to be a matter of active investigation. (Visual speech information is not the only type of nonauditory information listeners use. For example, gestures – a visual, but nonspeech cue – are also commonly used during conversational speech and tend to align with important acoustic cues.)

1.4 Major Themes

Throughout the remaining chapters in the book I will refer to a few major themes – central ideas that crop up in different contexts. This list is certainly not exhaustive, but it provides a starting point to think about important threads that run through different areas of language processing (and thus different chapters). By necessity, individual research studies need to investigate fairly specific questions, and the details of these studies are critically important to understanding the findings. At the same time, I hope these overarching themes encourage us to think about the big picture: What have we learned about language, or human cognition generally? One nice exercise is to return to these themes for reflection or discussion after each chapter: Although some will relate more obviously than others to each chapter, most chapters touch on several of these themes.

1.4.1 Stability versus Flexibility

Memory, in all its forms, is extraordinarily important for helping us to efficiently navigate our environment. When I was growing up, we had three cats, who got their dinner about five o'clock every afternoon. As soon as they heard the can opener, all three would come running so they wouldn't miss a morsel of food. Similarly, if you've ever had a bird feeder in the yard, you can see how quickly the neighborhood birds (and squirrels) learn that it provides a reliable source of food and they stop by to benefit. Remembering their past experience saves the animals from having to forage at random every time they're hungry, preserving that time and energy for other activities.

In some ways, language is no different. To efficiently communicate, we have collectively agreed upon a set of words that refer to particular concepts (and other words that help piece those together). If we each somehow invented our own personal language, it might be useful for helping organize our *internal* thoughts and plans. But the most obvious benefit of language comes from interpersonal communication: By developing a shared vocabulary, we can quickly communicate ideas (sometimes even very complex ideas) to others. As noted in Section 1.3.3, the need for a shared set of sounds we can use to communicate drives both the need for stability (we need to remember those specific sounds) and flexibility (we need to understand different talkers in different situations).

1.4.2 Language Processing Benefits from Prediction (or Context)

In the previous section I gave some examples of how speech comprehension is constrained by time, and thus how the rapid extraction of meaning is required during communication. Although there are no doubt a number of strategies listeners use to facilitate rapid language processing, many fall under the broad category of "prediction." That is, the average literate adult may know about 40,000 words,[10] but not all are equally likely to occur at any instant. Expectations about the next word to be heard will be influenced by the current social or environmental setting, preceding linguistic context, and acoustic cues in connected speech (among other things). The use of context to constrain perception – and specifically, to predict the upcoming speech signal – is fundamental to spoken language processing.

1.4.3 Language Processing Relies on Both "Bottom-Up" and "Top-Down" Processing

As typically used, **bottom-up processing** refers to sensory-driven computations carried out by the brain. For example, when photons of light hit the retina, a neural signal is created in the eye. The neural signal is completely explained by how the photoreceptors (light-sensitive cells) in the eye react to incoming sensory information. By contrast, **top-down processing** reflects non-sensory influences, including our expectations, memories, attentional state, and so on, that can alter the processing of incoming sensory information. Top-down influences are clearly demonstrated in the context of attention: If I provide a cue that indicates the spatial location of a flash of light, you will see it more rapidly than without the cue. The photons hitting your retina are identical, but your attentional state helps you to process that information more efficiently. In reality, the relative weight of bottom-up and top-down information is not always clear, and it may well be that dividing behavior into these two categories is an oversimplification.[11] Because of this ambiguity, I am tempted to always use quotes when describing these two positions – I will avoid doing so to keep the page cleaner, but you might want to imagine the quotes being there anyway. The degree to which our language understanding is driven by sensory information compared to other types of processing, and the degree to which this changes with context, is a recurring theme.

1.4.4 The Neural Organization of Language Processing Is Hierarchical

During everyday communication, language enters the brain through sensory regions that are also responsible for processing all of the nonlanguage things in our environment. That is, the auditory system that processes your friend saying "good morning" is also responsible for making sense of a dog barking, a car horn honking, and a running faucet from the next room. From here on, though, processing diverges, and increasingly complex forms of language processing are engaged, corresponding to different regions of the brain (particularly along the temporal lobe, and into the frontal lobe, covered in more detail in Chapter 3). The type of information processed at each stage changes and becomes less and less dependent on the original sensory input. In addition, the reciprocal wiring of the brain means that "higher" levels in the hierarchy can influence the processing at "lower" levels. The type of hierarchical organization I've described is hardly unique to language; rather, it seems to be a general organization principle of the brain. However, it plays a critical role in how we think about computations and constraints during language processing.

1.4.5 Language Processing Depends on the Task That People Are Doing

As listeners, it is critical that we can distinguish different speech sounds. For example, "bat" and "pat" are both valid words in English, with different meanings, and so we need to be able to tell them apart. Now let's say that I would like to conduct an experiment to study speech perception; I play a sound and ask listeners to indicate if they hear a "ba" or a "pa" (often referred to as a two-alternative forced-choice task, also abbreviated 2AFC). At first glance, this experiment might seem to transparently tap the same processes listeners use when understanding speech (distinguishing between a /b/ and a /p/ is required to distinguish between "bat" and "pat"). In reality, however, the task demands are very different. My experiment is using isolated speech sounds rather than words; moreover, I am asking participants to make an explicit decision about what they hear (that is a **metalinguistic** decision – a decision *about* language). Even though the decision might be an easy one, it still requires participants to consider options and make a conscious choice (and probably press a button to indicate the decision, requiring motor processing) – cognitive processes that are typically not present in everyday conversation. If I identify brain regions that respond to this sort of task, how sure am I that these are indeed what listeners do in the everyday world? Carefully controlled studies play an important and necessary role in helping us understand language. At the same time, we need to be aware of the specific task listeners are doing and the degree to which this affects our conclusions.

1.4.6 There Is No "Language Network"

Don't get me wrong – of course, some brain regions are more important for language function than others. Clinically, this is highlighted most obviously by the fact that only some types of brain damage lead to **aphasia** (severe language difficulty). However, this important (and undisputed) clinical observation, coupled with a general focus on localizing cognitive functions, has encouraged a mindset that often results in discussion of "the" language network, which is inaccurate. There is no single language region, language center, or even language network: We know beyond a doubt that the parts of the brain recruited to understand language are different depending on the