Introduction to the relation between speech comprehension and production

Niels O. Schiller and Antje S. Meyer

The present book explores how phonological and phonetic knowledge is represented and used in speech comprehension and production. These processes are obviously quite different: In one case a person hears and understands an acoustic input, whereas in the other case, the person produces an acoustic signal expressing a message to be conveyed. Therefore it is not surprising that many researchers have studied phonological and phonetic processing in either speech production or comprehension without considering the other process too much. This “divide-and-conquer” strategy has been very successful and has led to a wealth of empirical data and detailed production and comprehension models.

However, in addition to the obvious differences there are equally obvious similarities between speech production and comprehension: They both involve the use of linguistic knowledge; when a person learns a first or second language, productive and receptive skills develop hand-in-hand, and when language is impaired after brain damage, the loss is rarely confined to production or comprehension. When language is used in natural contexts, speaker and listener collaborate, as the speaker is the person creating the input for the listener and each speaker is also a listener of his or her own speech and, a moment earlier or later, the speech of the other person.

Given these obvious similarities and differences between speech production and comprehension, it seems worthwhile to ask how exactly they are related or different from each other. A number of specific research questions suggest themselves. For instance, is there one mental lexicon that holds all information required for production and comprehension and is accessed in both processes, or are there separate dedicated lexica for each function? If there are separate lexica,
are they organized according to similar or entirely different principles? How are the processes underlying speech planning and comprehension related? Are some processes involved both in production and comprehension? Are there processes that run in opposite directions in planning and comprehension? Which brain areas are involved in speech production and comprehension? Which are engaged in both and which in only one of these functions? These questions seem interesting and important to us in their own right. Providing answers to some of them should help us understand the overall architecture of the cognitive system involved in the use of language. In addition, evidence and theories concerning the relationship between production and comprehension should have implications for theories primarily concerned with production or with comprehension. Evidently, a theory of speech production should explain what is known about speaking, but it should be compatible with theories and evidence concerning speech comprehension. Likewise, a theory of comprehension should explain how people understand speech, but it should not clash with what is known about speech production.

In preparation of this book, we invited experts from different areas of psycholinguistics to discuss phonological and phonetic processing in speech production and comprehension. Unsurprisingly, the authors filled in this general brief in different ways.

Dell and Gordon discuss the effects of neighborhood density (the number words that are phonologically similar to a given word) in speech production and comprehension. Many properties of words have similar effects in production and comprehension tasks. However, neighborhood density is an interesting exception. High neighborhood density has detrimental effects in comprehension tasks, but facilitatory effects in speech production tasks. Dell and Gordon account for these results within an interactive activation model of speech processing. The opposing effects of neighborhood density arise because production and comprehension tasks create different competitive environments. Phonological neighbors are potent competitors to the targets to be selected in speech comprehension but not in production tasks, where the set of competitors is defined on semantic grounds.
McQueen, Dahan, and Cutler focus on the mapping of the speech signal onto stored representations in speech recognition and compare it to the analogical mapping of stored representations onto articulatory commands in speech production. They draw attention to striking differences between current comprehension and production models: All comprehension models assume that lexical access involves the parallel evaluation of several lexical hypotheses. This process is continuous, i.e. it does not involve discrete processing stages, and it is graded, i.e. lexical candidates may be more or less strongly activated depending on the amount of support they receive by the acoustic input. As the authors demonstrate there is strong empirical support for this view. By contrast, the evidence for continuous processing in speech production is far less convincing. Therefore, serial stage models of production co-exist with continuous models, which assume a limited degree of cascading of information. However, no production model posits the widespread cascading generally assumed in models of comprehension. In addition, the information passed on through the processing steps of production models appears to be rather more categorical than the fine-grained phonetic information used at all levels of the comprehension process. As McQueen and colleagues explain, these differences may be linked to differences in the tasks faced by language users during speech production and comprehension: A system continuously processing fine-grained phonetic detail is ideally suited to the purposes of speech comprehension, but such a system may not be perfect for production.

Zwitserlood discusses the representation and processing of morphological information. Drawing on evidence from studies of speech production and comprehension she presents a strong case for a separate level of morphological representation, distinct from syntactic, semantic, and phonological levels. Like Roelofs (though for different reasons) she argues that there must be separate phonological and phonetic components for production and comprehension, while the other levels, possibly including the morphological level, are likely to be shared.

Roelofs reviews several computationally implemented models of spoken word recognition and production and discusses whether a
single shared system can support word form access in production and comprehension. As he points out, such a system must have bi-directional links between sublexical and lexical units. This would imply feedback from the sublexical to the lexical level in production, and from the lexical to the sublexical level in comprehension. Roelofs reviews findings from a variety of sources and concludes that there is no evidence that would force us to assume such bi-directional links. He therefore proposes an architecture with separate but closely linked feed-forward systems for production and comprehension. This view is compatible with recent neuro-imaging studies suggesting that phonological processing in production and comprehension of speech engages separate subregions within Wernicke’s area.

Goldstein and Fowler discuss the relation between speech production and perception from the point of view of Articulatory Phonology. In Articulatory Phonology, vocal tract activity is analyzed into constriction actions (gestures) of distinct vocal organs (e.g., lip closure). These gestures are seen as atoms of a combinatorial system and are organized into temporally overlapping structures. The authors argue for the need of a "common currency" between production and perception and propose that gestures may serve as the common phonological currency. Gestures are preserved from language planning, via production to perception, and directly structure the acoustic signal. Listeners perceive gestures, not acoustic cues.

Guenther addresses the control of speech movements. In his view, speech motor planning and auditory speech processing are closely linked as, for instance, the targets for speech movements are defined in auditory and orosensory terms. Guenther proposes a neural network model of speech production and speech motor acquisition (DIVA), which captures speech-planning processes from the syllabic level to the level of muscle commands. (By contrast, Roelofs’ WEAVER++ model specifies speech-planning process up to the syllabic level). DIVA accounts for a wide range of behavioral data, including results of studies of articulatory kinematics and the development of speech motor skill development. It is also a model of the neural representations underlying speech production. Speech perception and production are supported by separate, but closely linked cor-
Van Turennout, Schmitt, and Hagoort describe how event-related brain potentials (ERPs), such as the N200 component, and lateraled readiness potentials (LRPs) can be used to determine the relative order in which speakers, listeners, and readers retrieve semantic, syntactic, and word form information and the time required for each of these processing steps. The elegance of this approach lies in the fact that very similar experimental paradigms and tasks can be used to study word production and comprehension. Van Turennout and colleagues review ERP studies showing that in picture naming, conceptual information is retrieved first, followed by the retrieval of syntactic information and, finally, the retrieval of the phonological form of the picture name. These results offer strong support for the WEAVER++ model of speech production (see the chapter by Roelofs), which predicts exactly this ordering of the retrieval processes. The model predicts that in word comprehension, the ordering of the retrieval processes should be reversed. The results of ERP studies (as well as other evidence) support the assumption that word form information is indeed accessed before syntactic and conceptual information (see also the chapter by McQueen et al.). However, the available data do not strongly constrain assumptions about the relative ordering of access to semantic and syntactic information.

Sebastián-Gallés and Kroll discuss the acquisition and organization of sound systems in bilingual infants and adults. An important theoretical issue they take up is whether the sound systems of the two languages of a bilingual speaker are strictly separated, or whether during the use of one language the sound system of the other language also becomes activated. They review the evidence from a large number of production and comprehension studies and conclude that lexical access is nonselective with regard to language. Interestingly, the task defines which part of the lexicon of the non-target language becomes activated. In production tasks, the activated names are those of translation equivalents of the targets, whereas in comprehension tasks, the activated words are words that sound similar to the targets.

Flege discusses the acquisition and representation of phonetic categories by second language (L2) learners and monolingual native
speakers of the same language. His Speech Learning Model (SLM) assumes that the cognitive capacities underlying speech acquisition remain intact in adults, i.e. even beyond a critical period for language acquisition postulated in some models. Adults retain the ability to form new phonetic categories for speech sounds encountered in a second language. However, native-like phonetic category formation becomes less likely with increasing age because the phonetic systems of the two languages are not completely separate. In the course of L1 phonetic category development (the end-point of which has not yet been determined) L2 speech sounds become more likely to be perceptually assimilated to existing L1 categories, blocking the formation of new L2 categories. Less accurate perception of L2 phonetic segments may lead to less accurate production of L2 speech sounds. Flege provides empirical evidence from production and perception studies demonstrating that (late) L2 learners indeed often do not distinguish to the standards of L1 speakers between different L2 segments.

As the reader will discover, the chapters of the book do not offer simple answers to the questions about the relationship between speech comprehension and production formulated above. Nor can they be joined together to form a unified theory accounting for speech production and comprehension. In fact, determining which production and comprehension architectures would be compatible is itself quite a complex issue. However, the reader will also discover that there is good agreement across authors on a number of important basic points: First, concerning the representation of phonological and phonetic knowledge, there is wide agreement that language users either have one shared store of knowledge which they access in different ways in speech production and comprehension, or separate, but closely linked stores. Nobody assumes that there are entirely independent representations used exclusively in production and comprehension. As pointed out by Roelofs, a shared system must be one with feedback between levels. This demonstrates that one’s view of the relationship between production and comprehension can have important implications for the design of the individual systems. A second point of consensus, highlighted in several chapters (Dell and Gordon, McQueen et al., and Zwitserlood) is that seemingly discrep-
ant findings from production and comprehension research can often be reconciled by carefully considering the task demands faced by listeners and speakers. Third, several chapters illustrate that similar experimental methods can be used for studying both domains, which is of utmost importance for systematic investigations of differences and similarities between speech production and comprehension. Similarly, computational models can be designed that can be used to simulate both production and comprehension processes. Finally, it is evident that cognitive models of speech processing should be developed in tandem with models of the neurological basis of these processes. We hope that these communalities represent a solid foundation for further research into the relation between speech production and comprehension.

We would like to end on a couple of personal notes. First, we wish to express our gratitude to the colleagues who reviewed chapters of the book for us. These are Dani Byrd, Albert Costa, Markus Damian, Laurie Feldman, Gareth Gaskell, Ray Kent, Tom Münte, Pascal Perrier, Rob Schreuder, Roel Smit, Marianne Starren, Patricia Tabossi, and Michael Vitevitch. Their contribution was greatly appreciated by the authors and the editors alike. Second, we wish to point out that this book is the result of a conspiracy by authors, reviewers and editors to create kind-of-a-Festschrift (only better) for a person who hates Festschriften but deserves one more than anyone else: Pim Levelt. Thank you, Pim, for everything!

Niels Schiller and Antje Meyer
Maastricht and Birmingham, March 2003