Libri

Niels O. Schiller, Antje S. Meyer (eds.)
Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities
Mouton de Gruyter,
Berlin 2003
355 pp.

The rationale for this volume is that much research in the past has focused either on speech perception or on production, with surprisingly little interaction between these two fields of investigation. The book contains nine chapters by expert researchers in perception and production, with the stipulated requirement that each contribution should make reference to both processes in order to highlight how they are similar and how they differ.

Of course, the processes are rather different, so it is not possible just to turn the speech production process backwards to understand how perception works. Listeners must make guesses about what is being said, so crucially they have to maintain and constantly re-evaluate parallel hypotheses on the basis of the incoming stream of sounds until they have determined which one is most likely to be correct, while production involves no such parallel guesswork. And speakers must achieve an extraordinarily intricate synchronisation of a whole range of different articulators, a muscular coordination effort which is not necessary for perception. However, presumably both processes access the same store of knowledge in the brain, so one assumes that the phonological representation of words must be in a form that both recognition and production can utilise. Furthermore, it seems likely that at least some of our speech handling faculties are shared by production and perception, so the question then is: at what level do the two processes become separate, and to what extent do they make use of shared utilities?

Some of the chapters in this book concentrate on the mental representation of phonology, while others focus more on the processes involved in speech production and perception, particularly considering models that attempt to simulate the processing of speech. One chapter that seeks to provide a summary of the current state of thinking in all these areas, including phonological representation as well as models of speech processing for both perception and production, is the one by James McQueen, Delphine Dahan and Anne Cutler. It consists of a substantial but rather densely packed overview of research into whether phonological information is categorical (dealing with phonemes) or more finely graded and whether phonological and lexical processing in perception and production proceeds in a serial or a cascaded manner. Their comprehensive summary concludes that, for recognition the processing is cascaded, with multiple candidate words being evaluated in parallel, and the information is finely graded, with access to far more detailed information than just phonemes, but in speech production the processing is mostly serial, with the possibility of only very limited cascading between the levels, and the information is categorical, dealing with phonemes and perhaps sometimes with an inventory of complete syllables. This fundamental contrast between perception and production makes sense, of course, as only in recognition is there a need for detailed goodness-of-fit scores to be maintained while multiple candidate parses are being evaluated in parallel, and moreover as semantic processing is sometimes essential for the successful identification of lexical items, strict serial processing would not work for speech comprehension.

However, the conclusion that simultaneous parallel processing occurs for the various stages during perception is challenged by Miranda van Turennout, Bernadette Schmitt, and Peter Hagoort, who report on the performance of WEAVER++ in modeling the temporal interaction of phonological, syntactic and semantic processing in the production and perception of speech, particularly in the light of electrophysiological research involving the lateralized readiness potential (LRP) and also the negative going potential N200 found in certain electrode sites near the front of the brain during a variety of picture-naming tasks. This neural imaging research indicates that, in speech production, there is mostly serial processing of information, from semantic concept to syntactic form and finally to phonological encoding, and this is well represented by WEAVER++. And even though this model was

originally designed for speech production, it can be adapted to deal with perception as well, in which case it predicts a reversed serial processing from sound through syntax to semantics. Indeed, N200 research on word recognition suggests that phonological processing precedes semantic processing by as much as 90 ms, which lends support to the predictions of the serial architecture of WEAVER++.

One possible explanation for the divergent claims of McQueen et al. and van Turennout et al. concerning whether the various stages of processing during perception proceed in parallel or not may lie in the tasks involved, as the investigations reported in the chapter on neural imaging involve only isolated words, whereas other research on comprehension often considers words in context, and it seems likely that there is a greater need for parallel processing of phonological, syntactic and semantic information when comprehension involves whole sentences.

Discussion of speech-processing models is central to many of the contributions in this book. The chapter by Ardi Roelofs provides an overview of some of the issues affecting the basic architecture, focusing on two models of speech recognition, TRACE and Shortlist, and two of production, DSMSG and WEAVER++. A fundamental contrast in the design lies in the approach to feedback, as TRACE and DSMSG incorporate feedback while Shortlist and WEAVER++ do not. For recognition, it is reported that Shortlist works perfectly well, and in fact there is no worsening in performance for TRACE when the feedback loop is disabled, so there seems to be no evidence for a need for feedback in speech recognition. And for speech production, WEAVER++ is able to model effects such as lexical biases (where speech errors involve words instead of non-words rather more often than expected) by means of a simple external feedback loop representing speakers' self-monitoring of their own production. Finally, Roelofs concurs with the findings of van Turennout et al. that WEAVER++ performs well for both perception and production in the light of data from neural imaging.

Another chapter that focuses on the architecture and performance of a speech-processing model is that by Frank Guenther, who provides an outline of the structure of DIVA, a model which attempts to represent the sensory motor skills required for the production of speech, especially the ability of infants to acquire those motor skills. One key issue is how humans can achieve motor equivalence immediately and effortlessly

even when a bite block is inserte d in the mouth, a phenomenon which DIVA is able to model quite successfully. It is interesting to note that such automatic compensation is assumed to be at least partially targeted towards achieving acoustic goals, not articulatory ones.

In contrast, when Louis Goldstein and Carol Fletcher consider the crucial issue of what kind of phonological representation is suitable both for production and perception, their framework is Articulatory Phonology, for which the representation of phonology is entirely in terms of articulatory gestures, denying a role for any kind of auditory or acoustic targets. This works well in modelling many speech production phenomena, especially speech errors which may be explained in terms of competing gestures, and it also accurately reflects evidence from experiments demonstrating that speakers can adopt compensatory mechanisms to achieve an appropriate gesture for /b/ or /z/ even when the natural movement of one of the articulators is artificially restricted. But what about perception? Do we really hear sounds in terms of articulatory gestures? And how can infants learn to associate articulatory gestures with the sounds they hear? This central issue is discussed in this chapter where it is reported that newborn infants can immediately mimic the action of an adult sticking out a tongue by attempting a similar gesture themselves, which provides evidence in favour of the direct mapping of what we perceive into physical gestures. Furthermore, in connection with the problem of mapping perceived speech onto gestures, it is reported that if infants as young as twelve weeks old are shown a film of a person mouthing a vowel at the same time as they hear a spoken vowel, they will look longer at a film where the vowels are the same than one where there is a mismatch, so it seems that there is indeed some evidence that infants can directly relate a perceived sound to a physical gesture.

However, apart from this brief consideration of the perception of vowels by infants, nearly all the material presented in the chapter on Articulatory Phonology relates to the production of consonants, and this overlooks the possibility that, while consonants may indeed be accurately represented in terms of articulatory gestures, the underlying nature of vowels may instead be based on auditory or acoustic targets, as there is compelling evidence that different speakers achieve perceptually the same vowel using radically different tongue positions [Ladefoged, 2005]. While Goldstein and Fletcher admit that many in the speech community

do not accept the claim that perception is in terms of articulatory gestures (p. 183), and they also acknowledge that many models, including that of Guenther (as discussed above), assume a role for acoustic as well as articulatory targets, they state that further discussion of this issue is beyond the scope of their paper (p. 182). This is perhaps a pity, as the question of acoustic versus articulatory targets does seem to be rather important in determining the phonological basis for speech processing, though this should not detract from the excellent presentation in this chapter of the basic tenets of Articulatory Phonology, including the useful discussion of how gestures might be learned by infants.

In most of the chapters that consider the interaction of various processes involved in production and perception of speech, little consideration is given to a clear role for morphology. However, the chapter by Pienie Zwitserlood does address the issue of how morphology affects speaking and listening, presenting the results of some fresh research investigating the interaction of morphology and syllable structure when Dutch listeners process words such as *helper* ('male helper') and *helpster* ('female helper') where the morpheme *help* spans two syllables in the former because of resyllabification but is fully contained within the initial syllable in the latter. On the basis of a monitoring experiment, it is shown that, while resyllabification does indeed occur with a derivational word such as *helper*, it is blocked in a compound word such as *melkemmer* ('milk bucket'), and on this basis, it is argued that morphological structure must play a part in the representation and processing of language.

In another chapter that focuses principally on speech processes, Gary Dell and Jean Gordon discuss the effects of phonological neighbourhood density, the degree to which a word shares similar pronunciation (differing by just one phoneme) with other words. Recent research has shown that high neighbourhood density benefits speech production but interferes with perception. Now, this makes logical sense, because a dense neighbourhood is one that is regularly accessed so production of the words is facilitated, while for comprehension, close neighbours are competitors that interfere with the correct identification of the word, but the challenge is for models of lexical-phonological access to represent the phenomenon accurately. Using an interactive model that seeks to replicate the errors of aphasic patients by manipulating the weights between the units and also the decay rate of activation, Dell and Gordon show that the effects of neighbourhood density on word production can be explained in terms of the feedback mechanism, the perceptual monitoring that takes place as we speak.

The last two chapters of the book consider speech processing and phonological representation from a fresh perspective, that of bilinguals and second language learners. If we remember that the majority of the people in the world are bilingual and furthermore that huge numbers of people are attempting to learn a second language, it should be obvious that processing two (or more) languages is actually the norm and not some kind of exotic skill mastered by a few talented individuals. Consequently, investigation of the skills involved in bilingualism and second language learning can provide fundamental insights into how language is processed, how phonological information is represented mentally, and also on issues regarding the plasticity of the brain.

Núria Sebastián-Gallés and Judith Kroll present a compact summary of recent research into the language perception and production of bilinguals, including late learners and adults learning a second language. Most of the research involves Catalan-Spanish, Dutch-English and English-French bilinguals, but there is also mention of other studies involving Hebrew-English, Arabic-French and Russian-English speakers. One conclusion that emerges from this wide range of investigations is that the human auditory system maintains a high level of plasticity, so we continue to be able to learn new sounds, but at the same time use of this ability requires substantial retraining which people often do not pursue, with the result for example that Spanish-Catalan bilinguals who are dominant in Spanish are generally not able to perceive vowel contrasts that exist for Catalan but not Spanish. A second conclusion of this chapter is that the two phonological systems of bilinguals are not separate, for even in tasks where only one language is clearly required, the second language continues to interfere with (or, in the case of cognates, facilitate) performance.

Finally, James Flege discusses the phonological processing of second language learners, specifically why it is that adult learners generally fail to achieve a native-like mastery of their second language phonology. He discusses two hypotheses for this failure: either the problem is connected with production, as speakers cannot learn to produce new sounds after the critical period; or it is perceptual, as adult learners cannot hear the L2 sounds properly. And there are two explanations for the perceptual hypothesis: either some of the

acoustic features for L2 sounds get filtered out; or the L2 features are wrongly interpreted, as their weightage gets warped by long years of exposure to L1. Research on the problems of /l/~/r/ for Japanese learners of English supports the warping explanation, as Japanese can hear the key features but they weight them differently from native speakers. However, as already mentioned in discussion of the chapter by Sebastián-Gallés and Kroll, evidence from Spanish learners of Catalan supports the filtering alternative, as many of these people simply cannot detect some of the features needed for accurate perception of their second language. The Speech Learning Model (SLM) developed by Flege and colleagues represents the warping hypothesis, as it assumes that the underlying capabilities of learners remain intact throughout their life, and even though perception may be affected by L1 exposure, native-like perception can be attained through careful retraining, which conflicts with the assertions of the critical period hypothesis. The SLM position is supported by research findings that many Italian and Spanish late learners of English can detect features such as rhoticity that are not important in their L1.

All of the chapters in this book provide valuable summaries of the state of the art regarding research into the processes and mental representations involved in speech perception and production, especially regarding computer models that represent these processes most accurately. While these summaries are all written by acknowledged experts from around the world and so are authoritative as well as interesting, especially with the new perspective that they consider both comprehen-sion and production, few of them report the detailed results of new research. Typical of the tone of most of the contributions is that of the chapters by McQueen et al. and also Sebastián-Gallés and Kroll, where the extent of the coverage of recent research is impressive and exceptionally valuable, but the fact that so much is included in about thirty pages means it is inevitably rather densely packed, and one ends up wishing that there might be a few charts to illustrate some of the issues, or some more substantial coverage of one or two of the research studies, or perhaps even the presentation of some original new research. A notable exception to this effort merely to summarise a wide range of recent research is the chapter by Zwitserlood which provides a breath of fresh air with its detailed description of an investigation into the interaction of phonology and morphology, but it is a pity that the absence of detailed reports on new, original research in much of the rest of the book leaves it somewhat dry.

Just occasionally, the final editing is a little lacking, for example with repeated material at the end of one page and start of the next one (pp. 50–51, 164–165, 288–289), missing fonts such as discussion of contrasts between '[t-T]' (p. 163) and '/d-D/' (p. 191) where one assumes that the second symbol in each case should be a dental fricative, strange symbol usage such as the suggestion of a back vowel in the first syllable of *panda* (p. 66), and other miscellaneous typos such as 'depends of preceding or following material' (p. 93), 'The shared the first morpheme' (p. 95), 'words hewere preceded by the other member of the minimal pair' (p. 293) and 'specified in advanced' (p. 305). Mostly these are merely irritating, though occasionally some material becomes a bit hard to understand fully, such as the reference to 'an /'bi/ disyllable' (p. 186) which is presumably missing a vowel at the start.

However, these are minor gripes and do not seriously undermine the value of what is by and large a handsomely presented book that provides a valuable overview of the field from a fresh perspective. While there is inevitably some disagreement between the contributions, particularly regarding simultaneous or serial processing of the stages involved in perception and also the question of whether phonological targets can all be represented in terms of articulation or should be based at least partly on acoustic or auditory features, the inclusion of these disagreements enhances the value of the book in providing an overview of issues that are being discussed in the field. In a brief review of investigations into the timing of speech processes done in the mid nineteenth century, Roelofs concludes that much recent research agrees closely with the original findings from the pioneering work of Donders. However, it is also true that the highly sophisticated range of new research techniques now available, especially those involving detailed neurological imaging, have thrown up a whole host of valuable new insights. Inevitably, these new techniques are still undergoing refinement, so it is not surprising that many of the findings are contradictory, and this book does a good job in summarising the current state of play in the field, including an overview of some of the controversies.

References

Ladefoged, P.: Speculations on the control of speech; in Hardcastle, Mackenzie Beck (eds.): *A figure of speech: A festschrift for John Laver* pp. 3–22 (Erlbaum Mahwah 2005).

David Deterding, Singapore

Schiller/Meyer