Phonotactic Constraints Shape Speech Perception: Implications for Sublexical and Lexical Processing

SEGUI, Juan, FRAUENFELDER, Ulrich Hans, HALLÉ, Pierre


Available at:
http://archive-ouverte.unige.ch/unige:85562

Disclaimer: layout of this document may differ from the published version.


---

11

Phonotactic Constraints Shape Speech Perception: Implications for Sublexical and Lexical Processing

Juan Segui, Ulricht Frauenfelder, and Pierre Hallé

The way in which we perceive speech is determined by the phonological structure of our native language. Though this conclusion may appear today as somewhat trivial, it is one of the more important advances in modern psycholinguistics, both from a theoretical and a methodological point of view. Indeed, language-specific specialization is now understood as necessary for humans to process their native language(s) with the fluidity and the fluency that have long been considered as a puzzling performance. Far from being a limitation, narrowing down speech perception to the native categories is required to cope with the variability inherent to the speech signal, and to benefit from "perceptual constancy." In recent decades, acquisition data have accumulated showing that language-specific specialization is largely guided by innate learning programs and is almost achieved by the end of the first year, both for speech perception and production. (This latter aspect has perhaps received less attention than it deserves.) It is interesting to note that Polivanov (1931) had already had this insight seventy years ago when he stated that "le phonème et les autres représentations phonologiques élémentaires de notre langue maternelle... se trouvent si étroitement liées avec notre activité perceptive que, même en percevant des mots (ou phrases) d'une langue avec un système phonologique tout différent, nous sommes enclins à décomposer ces mots en des représentations phonologiques propre à notre langue maternelle" [the phoneme and the other phonemic representations for our native language... are so intricately linked to our perceptual activity that even when we hear words (or sentences) from a language with an utterly different phonemic system, we tend to analyze these words in terms of our native language phonemic representations]. Other scholars
from this early period shared similar assumptions about cross-language perception (Troubetzkoy, 1939/1969; Sapir, 1921, 1939).

It follows from Polivanov's observations that researchers cannot understand how listeners process speech without referring to listeners' native language. Indeed, it is by relating the perception performance of listeners to the phonological structure of their language that considerable progress has been made. In the last three decades, researchers have been able to characterize the evolution of the perceptual capacities of infants as well as the speech processing mechanisms in adults. Jacques Mehler is one of the first psycholinguists to have understood the importance of establishing this relation. By working within this tradition, he has made major empirical contributions in research on both infant and adults.

This chapter presents some of our recent research that fits well into this tradition. It examines the role of language-specific phonotactic constraints in speech processing. The experiments provide a striking demonstration of a new kind of perceptual assimilation that depends upon the listeners' knowledge of the phonotactic constraints in their language. To set the stage for this presentation, we first summarize briefly some of the major findings that have shown the impact of a listener's phonological knowledge on language acquisition and use.

The Influence of the Native Phonological System on Speech Perception

Important findings in infant language literature have emerged from research on both the perception of individual segments and that of segment sequences. Young infants appear to be equipped at birth with "universal" capacities for discriminating speech segments. Thus, for example, it has been shown that infants were generally able to discriminate nonnative speech contrasts during their first months of life (Streeter, 1976; Trehub 1976). This capacity is declining by nine months or before. By ten to twelve months, infants have reorganized their perceptual system and can no longer discriminate consonant contrasts not present in the language they are learning (Werker and Tees, 1984). This loss of phonological contrasts that are not functional in the language environment reflects the emergence of phonological deafness as it is observed in adults. Examples of phonological deafness of adults to foreign contrasts are abundantly documented in the literature (e.g., Dupoux et al., 1997; Polka, 1991; Pallier, Bosch, and Sebastián-Gallés, 1997). It is important to note that adults are not deaf to all nonnative language contrasts: Their ability to discriminate nonnative language contrasts depends upon the articulatory-phonetic differences between native and nonnative phonemic categories. For example, Best, McRoberts, and Sithole (1988) observed that English-speaking listeners are able to discriminate Zulu contrasts between clicks, which are all quite distant from any English sound. The Perceptual Assimilation Model (PAM) proposed by Best (1994) explains the differential sensitivity to foreign language contrasts by appealing to the notion of phonological perceptual assimilation. According to this model, two sounds that can be assimilated to the same segment of the native language are difficult to discriminate. A good illustration of such perceptual assimilation can be found in the well-known case of Japanese listeners who have great difficulties in discriminating the American English /l/-/l/ contrast (Goto, 1971; Mochizuki, 1981). Indeed, Japanese listeners assimilate both English phonemes to the single Japanese phoneme /l/. Note that French listeners also have some trouble with the American English /l/, which they tend to assimilate to /w/ (Hallé, Best, and Levitt, 1999).

Another important aspect of infants' speech development is their increasing sensitivity to recurrent patterns in general (Safran, Aalin, and Newport, 1996), and more specifically to sound sequences or phonotactic regularities in the language they are acquiring. It has been shown that nine-month-old Dutch infants are sensitive to phonotactically legal on-set and offset clusters in Dutch words (Friederici and Wessels, 1993; see also Jusczyk et al., 1993). Moreover, Jusczyk, Luce, and Charles-Luce (1994) have shown that nine-month-old American infants prefer listening to monosyllabic nonwords with a high rather than a low phonotactic probability.

These data suggest that by the end of the first year, children have acquired the phonological knowledge that is relevant to the possible and likely combinations of sounds in their language. In the next section, we turn to the influence on adult speech perception of this kind of phonological knowledge, that is, the knowledge of phonotactic constraints.
The Role of Phonotactic Constraints in Adult Speech Perception

Polivanov claimed that the native language affects not only the way in which individual sounds are processed but also the way in which sequences of sounds are perceived. He illustrated this point by observing that when Japanese native speakers are presented with an English word like "drama," they say that they hear /do.ra.ma/. This latter sequence is consistent with the phonotactic rules of Japanese, which does not allow syllable-initial clusters. Here then, we have an example of a transformation of a sequence that is illegal into one that is permitted in the language. This phonotactic effect in Japanese has recently been studied experimentally (Dupoux et al., 1999). When Japanese participants were presented with pairs of nonsense words such as /ebzo/ and /ebuzo/, they failed to perceive the difference. They reported hearing the vowel /u/ in /ebzo/ even though there was not any acoustic correlate for this vowel in the signal. By contrast, French listeners did not hear /u/ in /ebzo/. The authors interpreted their results in terms of an extension of Best's PAM.

These findings and other similar ones suggest that listeners interpret a nonnative illegal input sequence by assimilating this sequence to sound patterns that exist in their own language. We shall call this transformation process phonotactic assimilation. Three types of phonotactic assimilation may be distinguished: (1) the listener "ignores" individual phonemes (or stress patterns: Dupoux et al., 1997) that are present in the signal; (2) the listener "perceives" illusory phonemes that have no acoustic correlate in the signal; (3) the listener "transforms" one phoneme in another.

The objective of this chapter is to present some recent findings that shed light on how phonotactically illegal sequences of native language sounds are assimilated to legal sequences according to the third type of assimilation. We will investigate the processing of /tl/ and /dl/ clusters that are illegal in French in word-initial position, although they are attested in this position in other languages. These sequences represent an accidental gap within the set of possible obstruent + liquid (OBLI) clusters in French. As we will show, /dl/ and /tl/ are perceptually assimilated to legal OBLI clusters by changing the place of articulation of the initial stop consonants. Our intention is first to demonstrate the existence of such phonotactic assimilation and, second, to uncover its time course and its perceptual cost. Finally, we examine the consequences of this phonotactic assimilation for lexical access. As it will appear, the outcome of these studies also bear on the general issue of how word forms are represented and how they are matched to the lexical input forms.

Experimental Studies

Massaro and Cohen (1983) conducted the first experiments on the perception of illegal consonant clusters in English. Here, participants were presented with synthetic speech stimuli beginning with an obstruent consonant followed by a "liquid" taken from an /r/-/l/ continuum. The category boundary in the /r/-/l/ continuum was shifted, relative to a neutral situation, such that more /r/s than /l/s were identified after /t/ and conversely, more /l/s than /r/s were identified after /s/. These results suggest the existence of a phonotactic assimilation process which biases listeners to perceive word-initial legal /tr/ and /sl/ rather than illegal /tl/ or /sr/. More recently, Pitt (1998) replicated these findings and further showed they were better explainable by phonotactic constraints than by cooccurrence frequencies.

In our first experiment, we wanted to discover what French listeners hear when they are presented with items beginning with the illegal cluster /dl/ or /tl/. A simple and straightforward approach to answering this question is to let participants freely transcribe what they hear when presented with such clusters. Nonwords with an initial /tl/ ("tlabdo," "tlobad," "tloab," and "tloba"), and with its voiced counterpart /dl/ ("dlapto," "dlapot," "dlapat," and "dlopta") were used as test stimuli. The outcome of this experiment was clear-cut: the /dl/ and /tl/ items were transcribed with a /gl/ or /kl/ cluster 85.4 percent of the time, and with a /dl/ or /tl/ cluster only 13.2 percent of the time. This finding was confirmed in a second experiment in which participants were given a forced choice identification task on the initial plosive consonant. Again, the results obtained showed a very substantial dental-to-velar shift for the items with an illegal cluster and suggested that some kind of phonotactic assimilation occurred for /dl/- and /tl/-initial items.
In order to eliminate alternative explanations for this shift in terms of the acoustic-phonetic properties of the illegal clusters, we conducted some acoustic analyses of the speech signal and performed two phonemic gating experiments. In this variant of the gating paradigm, participants are asked to transcribe exactly what they hear, rather than to guess words. The results showed that the initial stops of /dl/- and /tl/-items were first identified as being dental, not velar, in the early gates (first 60 ms after release burst). So it is unlikely that there were cues to velar articulation in the speech signal in the portion corresponding to the stop onset. This was confirmed by the acoustic analyses. However, at the point where the /l/ and the following vowel were first identified, subjects revised their dental responses and began to produce velar responses. This assimilation process for the initial stop consonant (dental to velar) thus depends upon the following phonological context. Hence rather than showing misperception of the illegal cluster's first constituent, these experiments provide a striking example of phonotactic assimilation.

Having demonstrated the existence of a phonotactic assimilation process in offline tasks, we then asked whether the process is also observed using a more online task such as the generalized phoneme monitoring task (Frauenfelder and Segui, 1989). The same /dl/- and /tl/-items as in the previous experiments were used. When the participants were instructed to detect /d/ and /t/ targets, they most often failed to do so (69 percent misses, on average). In contrast, when participants were instructed to detect velar targets in the same items, they did so (erroneously) 80 percent of the time. These findings suggest that listeners categorized the initial phoneme of the illegal sequence as a velar rather than a dental in real time. An additional finding of interest was the longer reaction times (RTs) to the assimilated target (/k/ in tlabod) than to the actual target (/k/ in klabod). Thus, we have shown that these illegal clusters induce phonotactic assimilation and require more processing time than legal clusters.

In more recent experiments, we explored whether the additional processing load associated with the phonotactic assimilation of the initial consonant cluster also affects the processing of later portions of the speech input. One way to measure later-occurring effects of a word-initial illegal cluster on subsequent processing is to have participants detect phonemes occurring later in the test items. In case the additional load induced by an illegal cluster is not confined to the processing of that cluster, phonemes that occur later in the speech sequence should be detected more slowly in illegal than in phonetically matched legal nonwords.

In this experiment, participants detected the target that occurred as the fourth phoneme in the item (e.g., /b/) either as the coda of the first syllable (/b/ in /tlobd/da/) or as the onset of the second syllable (/b/ in /tlobd/). This manipulation of the syllabic position of the target phoneme additionally allowed us to test whether the processing cost is confined to the syllable containing the illegal cluster or whether it propagates downstream to the initial phoneme of the second syllable.

The results showed that the presence of an illegal cluster in word-initial position slowed down the detection latencies to a phoneme located in either the first or the second syllable. Furthermore, the syllabic target position affected miss rate in illegal items: more misses occurred for phoneme targets in the first than in the second syllable. Therefore, the processing load due to the presence of a phonotactic violation appears to depend upon the syllabic position of the target.

It is possible that there were some acoustic-phonetic differences in the production of the target phonemes following legal and illegal clusters due to a difficulty the speaker may have encountered in producing the illegal stimuli. To control for this possibility, we conducted a further experiment in which the initial consonant-consonant-vowel (CCV) part of legal and illegal stimuli were cross-spliced. If the effect observed in the previous experiment was due to the presence of a phonotactic violation (rather than to reduced acoustic saliency of the target), then the same pattern of results should be observed. This was indeed the case. RTs were longer for the targets following illegal sequences. Moreover, both RT and miss rate data showed a significant effect of syllabic position, with faster and more precise detection of the target in the second than in the first syllable.

These experiments showed that detection of the word-medial stops was delayed by the presence of a word-initial illegal cluster /dl/ or /tl/. The increased difficulty in detecting the word-medial stops in illegal items was not due to phonetic-acoustic or prosodic differences compared to the legal items. It is important to note that the greater processing cost observed in the detection of the coda of the first syllable has a structural rather
than a temporal origin since in both conditions being compared the target phoneme always occurred in the same serial position. These results suggest that the syllable plays a role in the perceptual integration process (Hallé, Segui, and Frauenfelder, in preparation).

The final series of experiments use lexical decision and cross-modal priming paradigms (Hallé, Segui, and Frauenfelder, submitted) to evaluate the effect of the phonotactic assimilation process at the lexical level. In particular, an auditory lexical decision experiment compared the responses to nonwords which either could or could not be assimilated to existent words. For instance, a nonword like “dlaicel”—beginning with the illegal consonant cluster /dl/—can be assimilated to the word “glæuel,” whereas the legal nonword “droleel”—beginning with the legal consonant cluster /dr/—should not be assimilated to the word “groseel.” Note that these two nonwords differ from the base words from which they are derived by the same, single phonetic feature (namely, dental vs. velar place of articulation), and thus are equally distant phonologically from their respective base words. The results showed that more “word” responses were observed for the phonotactically assimilated items (68.3 percent) than for the legal nonwords (19.6 percent), which served as a control condition.

A more subtle test was performed using the cross-modal repetition priming paradigm. Here participants heard either the base words or the derived nonwords (“dlaicel” or “droleel”) and made a lexical decision on the base words, presented visually (“glæuel” or “groseel”). The findings indicated that RTs to the words like “glæuel” were similar whether the prime was the real word (repetition condition) or the phonotactically assimilated nonword (471 ms vs. 496 ms). In contrast, a large RT difference was obtained for the case of phonotactically nonassimilated nonwords (469 ms vs. 552 ms). Taken together, these experimental findings demonstrate the effect of phonotactic assimilation at the lexical level. In addition, they also suggest that the amount of lexical preactivation produced by the prior presentation of a phonologically related nonword does not depend solely upon the phonological distance between prime and target measured in terms of distinctive features but is also determined by the phonotactic properties of the prime. Indeed, given that prime-target pairs always differed by the same phonological distance, the differential amount of lexical preactivation observed can only be related to the legal vs. illegal nature of the initial cluster.

Discussion

In this chapter we have presented the main findings of a series of experiments that explored a case of phonotactic assimilation. We have demonstrated a simple yet robust phenomenon: When listeners are presented with the illegal consonant cluster /dl/ or /tl/ they report hearing the legal cluster /gl/ or /kl/, respectively. The observed effect seems to be very robust and unrelated to the acoustic properties of the signal. In fact, a gating experiment showed that listeners initially perceived the first consonant of the cluster correctly and only somewhat later, when they were able to identify the second consonant of the cluster, began to revise their responses and report the phonotactically legal cluster /gl/ or /kl/.

Experiments using a phoneme monitoring paradigm confirm the online nature of the assimilation process and furthermore reveal a processing cost associated with the phonotactic assimilation. This processing cost was observed not only in the detection latencies of the initial member of the illegal cluster but also for phoneme targets located at later positions in the carrier items. Moreover, the size of the perceptual cost as reflected by both miss rates and RTs was greater for phonemes belonging to the syllable containing the illegal cluster than for phonemes located in the following syllable.

Finally, we studied the effect of this type of phonotactic violation in lexical access. Phonotactic assimilation was found to operate for words just like it does for nonwords, so that altered forms obtained by changing, for example, /gl/ to /dl/, are readily mapped to the lexical forms from which they are derived. We return to this point later.

Overall, the experimental results presented in this chapter confirm the role of phonotactic constraints in speech perception. We interpreted the results in terms of an extended version of PAM according to which an illegal sequence phonologically similar to a legal sequence is assimilated to that legal sequence. Differently said, listeners hallucinate another phoneme than that which actually occurs in the sequence, illustrating the third type of phonotactic assimilation described above. It is important to
note that these experiments compared the perceptual processing of legal and illegal items within the same language. Therefore, the case of perceptual assimilation we found is not explicitly addressed in PAM, which in its current version focuses on cross-language perception. How could PAM deal with the case of within-language perceptual assimilation that we found? PAM makes predictions about the type of assimilation pattern and the level of discrimination performance that should be observed for a nonnative contrast, according to both the phonetic and structural relations between native and nonnative phonemic categories. Yet, we can consider that /dl/ and /tl/ are foreign to the phonology of French. In the terminology of Best's model, the fact that French listeners hear /dl/ as /gl/ is most readily interpreted as a case of category goodness (CG) assimilation: /gl/ is indeed a good exemplar of /gl/, while /dl/ is probably heard as an atypical exemplar of /gl/. (The reason why /dl/ assimilates to /gl/ rather than to /bl/ is to be found in the phonetic-acoustic properties of the sounds involved; see Hallé et al., 1998.) That the phonotactic assimilation we found is related to the CG type rather than, for example, the single category (SC) type of assimilation could be tested crosslinguistically using categorization and discrimination tasks. For example, modern Hebrew allows both velar and dental stop + /l/ clusters. Given the abundance of evidence for crosslinguistic deafnesses, we expect that French listeners would have difficulty in discriminating the Hebrew /dl/-/gl/ and /kl/-/tl/ contrasts. More specifically, we predict that French listeners would categorize the dental stops of these clusters as velar.

In addition to contributing to a more complete account of perceptual assimilation in speech perception, our data also shed some light on word recognition. Indeed, the pattern of data obtained in the cross-modal experiment is informative about the mapping process underlying word recognition. It is generally assumed that a lexical item is recognized by means of a mapping from the speech signal onto its representation. The activation level of any given lexical entry is mainly a function of its “vertical similarity” to the input (Connine, 1994; Connine, Blasko, and Titone, 1993), that is, the goodness of fit between the speech input and a particular lexical representation. (The activation level also depends on the possible competition with phonological neighbors: this modulates what Connine, 1994, called “horizontal similarity.”) Previous experiments indicated that vertical similarity takes into account two main factors: the amount of deviation between input and lexical forms in terms of the number of mismatching distinctive features, and the serial position of the mismatching phonemes (see, e.g., Marslen-Wilson and Zwitserlood, 1989). The results obtained in our research showed that the goodness of fit also depends on the phonological context immediately following the altered phoneme, and in particular, on the possibility that this context induces “phonotactic assimilation.” In terms of phonetic features and serial position, “dlaïeul” differs from “glæïeul” in the same way as “droseïeul” differs from “groseïeul.” Yet, only the former nonword strongly activates the related word. In other words, goodness of fit must also take into account contextual similarity and language-specific phonotactic knowledge.

To the extent that our cross-modal repetition priming data reflect real-time processes, the differential priming effects for “dlaïeul” vs. “droseïeul” suggest that lexical access proceeds through a stage of perceptual integration into sublexical units larger than phonemes or phonetic features. Our results are not compatible with the notion of a direct mapping of a sequence of phonemes onto lexical forms. A mapping from a structured array of phonetic features extracted from the input is equally unlikely for the same reasons. (Also, such a mapping is not consistent with the processing discontinuities observed in the phonemic gating data at a sublexical level.) Our data thus point to the mandatory perceptual integration of the speech flow into sublexical units that are at least of the size of complex syllable onsets. Speech seems to be processed with an extreme ease and fluidity when it is free from possible distortions, and this might suggest a continuous, direct kind of processing which can be based on very fine-grained units of perception. Our data, on the other hand, strongly suggest that speech sounds also are integrated into coarser-grained units such as syllable (complex) onsets. In our view, these two notions are not necessarily contradictory. We suggest that different analyses of the speech input are performed in parallel, and, as it were, compete to produce a plausible outcome as quickly as possible. We therefore do not claim a precise size for the coarse-grained units which were evidenced in our data. Could they be the size of the plain syllable? Although this is prima facie not in line with recent propositions that dismiss the syllable
as a processing unit (Cutler et al., chapter 10 in this book; Frauenfelder and Content, 1999), our findings of a greater sensitivity to the processing load induced by phonotactic violations within than across syllables suggest that the syllable still is a good candidate for a prelexical unit of speech integration.

Acknowledgments

This research was supported by a grant from the Fonds National de la Recherche Suisse (FNRS 1113–049698.96) and a BQR grant from the René Descartes University, Paris V.

References


Research by Jacques Mehler and his collaborators (e.g., Cutler et al., 1989, 1992; Mehler et al., 1993, 1996) has shown that the processing routines (prelexical segmentation, word segmentation, etc.) that are engaged in speech perception are not identical from one language to the other, but are finely tuned to the specific properties of the native language. In contrast, current models of speech production have emphasized the language-universal aspects of the process; that is, the details of the processing routines (processing levels, units, time course of the computations) are always thought to be identical across languages. Is speech production special in that it uses exclusively language-universal procedures, or is there as much language-specific tuning in speech production as has been found in speech perception? Here, we argue that the cross-linguistic investigation of noun phrase (NP) production reveals language-specific differences in the procedures used to select determiners and other closed-class words.

There is a clear difference in the kind of information that is used to select open- and closed-class words in language production. The selection of open-class words, such as nouns and verbs, depends primarily on their individual meanings. For example, in sentence 1a, the selection of “rose” depends only on its meaning (to fulfill the speaker’s intention to communicate the proposition: is on rose, table). Quite a different process seems to be at play in the selection of closed-class words, such as determiners and pronouns. The selection of the latter types of words depends largely on properties of other words in the sentence. This is clearly illustrated in sentences 1b–d for determiners and in sentences 1e–g for demonstratives and possessives.