COARTICULATORY RESISTANCE IN A MENTAL SYLLABARY

Uta Benner, Ines Flechsig, Grzegorz Dogil, Bernd Möbius

Institute of Natural Language Processing, University of Stuttgart
`{uta.benner;ines.flechsig;grzegorz.dogil;bernd.moebius}@ims.uni-stuttgart.de`

ABSTRACT

In a speech production model proposed by Levelt a distinction is made between two routes of phonetic implementation in speech. A syllabary route is used to retrieve the stored motor programs for the most frequent syllables of a language, and segment-by-segment assembly is used for the implementation of low-frequency syllables. One of the predictions of the model is that there should be a difference in coarticulation between motor programs retrieved from the syllabary and programs that are computed online. In this paper we present two laboratory experiments and a corpus study on German which were designed to verify this prediction. Our results support the hypothesis that articulatory programs for high-frequency syllables are implemented differently than those for rare syllables.

Keywords: speech production, coarticulation

1. INTRODUCTION

Corpus studies in Germanic languages for which large annotated speech databases are available (e.g., English, German, Dutch) show that about 500 syllable types (out of many thousand phonologically possible syllable types in these languages) cover over 80% of the corpus [6]. A small number of syllables is used extremely frequently and a majority of possible syllables occur in large corpora rarely, once, or not at all [11]. The appealing idea that the motor programs for these extremely frequent speech events should be stored in a mental repository—the syllabary—which in turn can be independently accessed in the process of speech production has found support in a number of psycholinguistic, neurolinguistic and phonetic studies.

The concept of the syllabary was introduced as an elegant explanation of some of the slips of the tongue speech errors [4]. The implications of this concept for reaction time scores of high-frequency and low-frequency syllables were investigated in a number of experimental studies carried out at MPI Nijmegen. Early support for the hypothesis provided by [7] could not be verified in more carefully designed studies [9]. However, recent results from the MPI using the experimental paradigms of priming and symbolic learning provide full support for the concept of syllabary storage and for a dual route of speech production [3].

Further support for the syllabary was provided in an influential clinical phonetic study [16]. It was argued that patients suffering form apraxia of speech (AOS) have lost direct access to the overlearned motor patterns of very frequent syllables. The subsyllabic route model of apraxia of speech advanced by [16] predicts that AOS speakers will be disfluent, but it does not predict the error on the segmental level. The model also predicts that the error patterns would be blind to syllable structure.

In a carefully designed study [1] tested the predictions of the subsyllabic route model of AOS. They showed that all of their ten subjects showed error patterns that were indeed sensitive to syllable structure. It was shown that segmental clusters at the contact of two syllables are treated by the patients differently from clusters within the syllable. The result is incompatible with the subsyllabic route model of apraxia of speech advanced by [16].

The influence of syllable frequency on error rates was also investigated [1]. All apraxic patients produced significantly fewer errors for syllables ranking among the 100 most frequent syllables of German. This result indicates that AOS patients do have access to a mental syllabary.

A similar conclusion can be drawn from the very first neuroimaging study which explicitly investigated the role of a syllabary in speech production [8]. This fMRI study, in which subjects read German nonsense words constructed from syllables of very high and very low frequency, respectively, elicited additional activation patterns in the left temporal cortex (BA 20 and additional foci in BA 36, 37) for very high-frequency syllables (the syllabary). These activations, however, were always accompanied by the activation of the motor and the premotor left cortices, characteristic for online segmental assembly. Hence, the mental syllabary route appears to be used in parallel to the segmental assembly route.

Experimental phonetic studies of the mental syllabary are rare, and the evidence that they provide is ambivalent. [5] investigated the influence of the syllabary on the durational and coarticulatory patterns of frequent and rare English syllables embedded in nonsense words. The coarticulation data provided some support for the claim that speakers use different coarticulatory procedures for high-frequency syllables, but there were no differences in the syllable duration data.

Evidence for different duration patterns of syllabary and non-syllabary units was provided by [15] in an investigation of a German speech corpus comprising a very large number of syllables and segments [14]. They argued that the durations of syllabary units are stored as temporal patterns for the whole syllable, whereas the duration of an infrequent syllable is computed from the concatenation of segmental durations. A better fit was obtained for the regression of z-scores of the durations of syllables against the durations of the constituent segments in non-syllabary units vs. syllabary units.

In summary, psycholinguistic, clinical, and phonetic evidence does not present a coherent picture of the role played by the syllabary in speech production. There are apparently alternative efficient computations that can lead to highly coarticulated motor programs. The following three experiments further investigate the coarticulatory processes in laboratory speech and in acted speech.

2. EXPERIMENT 1

2.1. Procedure

A list of 400 bisyllabic nonsense words adapted from [8] was used in the experiment. Half of the words were constructed each by selecting two syllables from the set of the 1000 most frequent syllables in German, the other half contained syllables from the set of the 1000 rarest syllables. Syllable frequency information was based on syllable probabilities induced from multivariate clustering [12] using the CELEX database, which allows an estimation of the theoretical probability even for unseen syllables. In [12] probabilities were computed for approximately 40,000 German syllable types, classified according to the five dimensions of onset, nucleus, coda, position in word, and syllabic stress.

Only complex syllables of the types CVCC and CCVC were used to compile the test word list. This procedure generated possible, unattested words of high phonotactical complexity (e.g., “Korrtgant” as a possible high-frequency word, “Börsgesk” as a possible low-frequency word).

Sixteen native German subjects (8 m, 8 f, 20–35 yrs.) produced the words embedded in the carrier phrase “Ich habe *Börsgesk* gesagt.” The subjects read the sentences from the computer screen in an anechoic chamber. Slips of the tongue and other disfluencies were monitored online.

2.2. Results

The ratio of slips of the tongue was computed. The utterances were automatically annotated on the segmental level by forced alignment and vowel boundaries were manually checked and, if necessary, corrected. Formant values were automatically extracted using the ESPS *formant* program. F1 and F2 values were normalized by a standard method [10], which transforms the raw acoustic data into a speaker-independent perceptual vowel space. Statistical analysis (t-test on average) of the parameters of the perceptual vowel space was performed using the R package.

We found a high ratio of production errors (27%). Errors occurred more often in infrequent (58%) than in frequent syllables (42%). However, the total number of slips of the tongue suggests that the complexity of consonant clusters as well as possibly the difficulty of grapheme-to-phoneme conversion rather than syllable frequency was the main source of error [1].

Because formant structure is affected by coarticulation, formant values were x1=log(F1/SR) and x2=log(F2/F1) [10] which define vowels in the perceptual space. The t-test showed significant ($p<0.05$) differences between frequent and infrequent syllables for these two parameters (e.g., ɔ, o, ʊ, a). Yet for some vowels the difference did not reach significance in parameter x1 (e.g., ɛ, ɪ) or x2 (e.g., e, u).

3. EXPERIMENT 2

3.1. Procedure

The high frequency of production errors in Experiment 1 has motivated us to test the degree of coarticulation on two new lists of word tokens with a simpler syllable structure. The first list consisted of bisyllabic nonsense words in which the initial syllable systematically varied in terms of frequency as in Experiment 1 but had a simple CV or CCV structure. The second syllable had a constant phonological form for all word tokens (e.g., high-frequency “gietet”, low-frequency “grehtet”). The second list was adapted from [1] and consisted of real German words subdivided into high-frequency and low-frequency initial syllables (e.g., “Diele”, frequency rank of /diː/ = 1 vs. “Dose”, frequency rank of /doː/ = 9512). These lists were produced by the same speakers and using the same procedure as in Experiment 1.

3.2. Results

There were significantly fewer production errors (4%) by the speakers for the simple words list than for the list in Experiment 1. The distribution of errors between frequent and infrequent nonwords was the same (42% vs. 58%) as in Experiment 1. There were no errors in the production of the real words.

The variables x1 and x2 were found to be significantly ($p<0.05$) different between frequent and infrequent nonwords for all vowels except /e/ (only x1 significant) and /aɪ/ (neither x1 nor x2 significant). The picture was less clear for real words where for some vowels there was no significant difference at all (e.g., /œ/). For other vowels only one value was significantly different (e.g., /ʊ/ and /a/).

4. EXPERIMENT 3

4.1. Procedure

In this experiment coarticulatory effects on vowels were investigated by use of minimal syllable pairs extracted from a large acted speech corpus comprising 17,489 words, 34,000 syllables, 2601 sentences recorded by a male speaker [14]. The corpus was designed to optimize the coverage of features and their combinations in language and speech. It was annotated on the segmental, syllable, word, part-of-speech, and intonation levels.

Minimal syllable pairs were obtained as follows. Each syllable in the corpus was annotated for frequency based on the same probabilistic syllable classes as in Experiment 1. Syllable pairs were formed from segmentally identical syllables in identical positions in the word, differing only in terms of syllabic stress status (e.g., /ˈma:/ vs. /ma:/) to make the pair partners as similar as possible. The mean of the probability values was used as a threshold for separating high- from low-frequency syllables within an identical pair [7]. Syllables with a frequency of occurrence of less than five were excluded from the study. 30 minimal syllable pairs were available for the analysis.

Formants were extracted using the ESPS *formant* program. The means of the frequencies of the first three formants were computed at nine equidistant steps during the duration of the syllable over all occurrences of a minimal pair member.

4.2. Results

Large differences were found between frequent and rare syllables in formant values, trajectories during the syllables, and transitions at syllable boundaries.

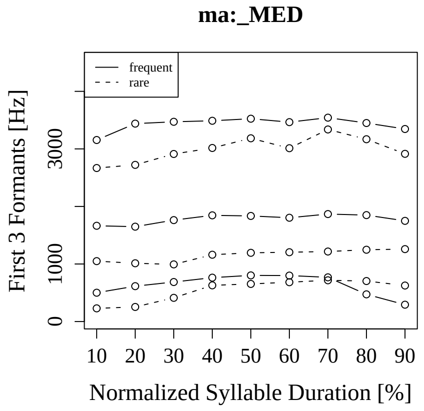

Figure 1: Formant trajectories in the syllable pair /maː/ in word-medial position. Solid lines represent averaged formants in the frequent syllable tokens, dotted lines represent those in rare ones.

The differences can be ascribed mainly to coarticulatory effects. Figure 1 is representative of the results. For instance, the difference in the F2 trajectories exhibits a target undershoot in frequent syllables. The frequent tokens of the /maː/ syllable pair had a mean duration of 137 ms, whereas the rare tokens had a mean duration of 200 ms. Thus the high-frequency syllables are more prone to coarticulatory effects. The differences in the F1 and F2 trajectories indicate that the frequent tokens of /maː/ are produced with a more centralized vowel than the infrequent ones. Centralized vowels arise from more “neutral” lingual gestures and are less resistant to coarticulation than “clear” vowels. Furthermore, faster formant transitions were observed in frequent syllables, a consequence of their special coarticulatory behavior, e.g., in the rising transition of F1 and F3 at the syllable start (see Figure 1).

Duration differences can of course occur in pairs of segmentally identical syllables as a function of the position of the syllable in the word or of syllabic stress or both. However, the differences observed in the formant trajectories in the 30 syllable pairs were consistent only with respect to syllable type frequency. In some of the pairs the frequent syllables were the stressed ones, in other pairs the frequent tokens were the unstressed ones. It is also worth noting that syllabic stress affects different classes of phones differently. Due to the overriding constraint to find suitable pairs of syllables according to frequency bins, the variation in syllabic stress was not optimally controlled for. The effect of the contrast in position-in-word was negligible.

The stronger coarticulatory effects found in the frequent syllables compared to the infrequent counterparts may be ascribed to the existence of a mental syllabary. Other theories of speech production can hardly account for such differences in the degree of coarticulatory resistance in pairs of syllables that are segmentally identical and differ just in their frequencies of occurrence.

5. CONCLUSION

Three experiments were carried out to test the prediction that coarticulatory resistance is stronger in syllables with a low frequency of occurrence, because rare syllables are assembled during speech production from the specifications of their constituent segments. In contrast, we expected a stronger degree of coarticulation in syllables with a high frequency of occurrence because, according to the concept of a mental syllabary, the articulatory programs of high-frequency syllables are stored for the entire syllable. Coarticulatory effects such as lenition due to frequent productions [13] are therefore expected to affect syllabary units more strongly than rare syllables.

These predictions were largely borne out. In Experiment 1 significant differences between frequent and infrequent syllables in pseudowords were found for two major parameters which define vowels in the perceptual space. However, due to high syllable complexity, a large number of production errors were observed. A simpler syllable structure in pseudowords and real words was used in Experiment 2. Again, the perceptual vowel space parameters were found to be significantly different between frequent and infrequent nonwords. The effect was also present, but less consistent, in real words. Finally, in Experiment 3 coarticulatory effects on vowels in real words produced by a single speaker were investigated. Significant differences were found between frequent and rare syllables in terms of overall formant values, formant trajectories during the syllables, and formant transitions at syllable boundaries.

Overall, we found a tendency towards stronger coarticulation and greater coarticulatory variability in high-frequency syllables than in low-frequency ones. This effect was first observed in laboratory speech and subsequently validated and confirmed by investigating it in a larger corpus. Our findings are compatible with the concept of the retrieval of high-frequency syllable-sized units from a syllabary and with the assumption that syllabary units are less resistant to coarticulation than syllables assembled from segmental specifications. However, retrieval from a syllabary is not necessarily a more efficient computation. Speakers have a deep phonotactic and statistical knowledge about segments, segment sequences, and syllables. The segmental assembly route is probably always available even for the production of frequent syllables, for instance to produce “clear” speech.

6. REFERENCES

[1] Aichert, I., Ziegler, W. 2004. Syllable frequency and syllable structure in apraxia of speech. *Brain and Language* 88, 148–159.

[2] Baayen, Harald and Piepenbrock, R. and Gulikers, L. 1995. The CELEX lexical database—Release 2.

[3] Cholin, J., Schiller, N., Levelt, W. 2004. The preparation of syllables in speech production. *J. Memory and Language* 50, 47–61.

[4] Crompton, A. 1982. Syllables and segments in speech production. In: Cutler, A. (ed), *Slips of the Tongue and Language Production*. Berlin: Mouton, 109–162.

[5] Croot, K., Rastle, K. 2004. Is there a syllabary containing stored articulatory plans for speech production in English? *Proc. 10th Austral. Internat. Conf. Speech Sci. Tech.* Sydney.

[6] Levelt, W. 1999. Producing spoken language: a blueprint of the speaker. In: Brown, C., Hagoort, P. (eds), *The Neurocognition of Language*. Oxford UP, 83–122.

[7] Levelt, W., Wheeldon, L. 1994. Do speakers have access to a mental syllabary? *Cognition* 50, 239–269.

[8] Mayer, J., Ackermann, H., Dogil, G., Erb, M., Grodd, W. 2003. Syllable retrieval vs. online assembly: fMRI examination of the syllabary. *Proc. 15th ICPhS* Barcelona, 2541–2544.

[9] Meyer, A. 1992. Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition* 42, 181–211.

[10] Miller, J.D. 1989. Auditory-perceptual interpretation of the vowel. *JASA* 85(5), 2114–2134.

[11] Möbius, B. 2003. Rare events and closed domains: Two delicate concepts in speech synthesis. *Internat. J. Speech Tech.* 6(1), 57–71.

[12] Müller, K., Möbius, B., Prescher, D. 2000. Inducing probabilistic syllable classes using multivariate clustering. *Proc. 38th ACL Meeting* Hong Kong, 225–232.

[13] Pierrehumbert, J. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In: Bybee, J., Hopper, P. (eds), *Frequency and the Emergence of Linguistic Structure*. Benjamins, 137–157.

[14] Schweitzer, A., Braunschweiler, N., Dogil, G., Klankert, T., Möbius, B., Möhler, G., Morais, E., Säuberlich, B., Thomae, M. 2004. Multimodal speech synthesis. In: Wahlster, W. (ed), *SmartKom: Foundations of Multimodal Dialogue Systems*. Springer, 411–435.

[15] Schweitzer, A., Möbius, B. 2004. Exemplar-based production of prosody: Evidence from segment and syllable durations. *Speech Prosody 2004* Nara, 459–462.

[16] Varley, R., Whiteside, S. 2001. What is the underlying impairment in acquired apraxia of speech? *Aphasiology* 15, 39–49.