An Inferential Phonological Connectionist Approach to the Perception of English-Assimilated Speech

Hiba Zaidi
Department of Foreign Languages
Yahia-Fares University
Medea, Algeria
hibato92@hotmail.com

Abstract

This research is an attempt to investigate the perceptual system processing of English connected speech. More specifically, it attempts to offer an experimental study to show the impact of the phonological context and specific English learning experience on non-native listeners perception of a set of assimilated-English speech forms inserted in pair of words and sentential contexts. Given that assimilation of place as a regular phonological process shaping English casual speech, seems to pose a notorious problem for the perceptual system in general and for non-native listeners to English continuous speech in particular. This triggers the current research investigation of a networks performance on English assimilated forms so as to have rather measurable and testable results.

1 Introduction

Connectionism conventions of human cognitive phenomena, according to (Rogers, 2009), appeal for a propagated activation through interconnected units of the system. In this realm, speech processing area (speech perception) has caught a large area of computational connectionist models interest. Taking that speech perception is organized around a lexical access process (Traxler, 2012), (JEFFREY and JAMES, 1988), thus, lexical mismatch intolerance is highly invoked while phonological variations in English speech production are widely attested. More pointedly, assimilation of place is regarded as a regular process shaping English connected speech, and is disruptive to the recognition system function .i.e., it interrupts the lexical access process through which speech inputs are matched to their lexical representational units stored in the long-term memory. Consequently, a misperception and sometimes lexical ambiguity of the speech arise, as argued by (MecQueen et al., 1999). It incurs a mismatch between the incoming variant segments and their lexical representational entries. This prompts the question as how the system handles such perceptual perturbation Inferential Phonological Connectionist Model (Gaskell and Gareth, 2003) (Mg and Wd., 1996a) (Mg and Wd., 1996b) (Marslen, 1998) seems to be central for addressing place-assimilated speech drawing on two main notions. First, the perceptual system processes the assimilated forms through a gradual and regressive mapping of surface (output unit) onto underlying (input unit) by making use of SNR architecture (Mg and Wd., 1996a). Second, the system is very sensitive to the phonological conditions surrounding the assimilated forms and tends to infer the phonological rules regulating such variations. However, driving on the expectation that the network might have an active role in shaping the phonological rules regulating the assimilation process rather than simply inferring them, this incites the current research to approach the issue concerned based on a modified SNR version . In light of the aforementioned predictions, this research will proffer an experimental study on a network performance on a set of assimilated forms embedded in pair words and sentential contexts. The research will draw from the experimental design to guide the different empirical phases, including the preliminary pre-test, treatment and post-test phases.

2 Phonological Inferential connectionist Model

Phonological Inferential Model (Gaskell and Gareth, 2003) (Mg and Wd., 1996a) (Mg and Wd., 1996b) (Marslen, 1998) is a computational approach directed to speech processing area. It has been developed to find conceivable answers to recurrent questions faced in cognitive theories of speech perception, as how speech is processed. The model evaluates the interconnection between a set of phonological rules and given phonological contexts of a speech stream, in the sense that, as (Mg and Wd., 1996b) pointed, these rules are context-sensitive as they specify the context they can apply to (p.288). For example, the likely occurrence of / k/ as a variant of the lexical /t/ in the word thatking to be realized as /kkI/ at the surface level, is permissible only if the following segment has a velar place feature like /k/, providing a source for a velar place feature to be repressively (leftward) assimilated. Given the crucial property of phonological rules as being phonological context-sensitive rules, this amount to characterize the model for a heuristic account for speech perception for the reason that it calls in simultaneously both listeners specific phonological experience (acquired phonological rules) alongside the range of phonological context surrounded-variation. As far as speech processing area is concerned, the current model (IPM) holds for three main assumptions: first, a successful completion of the lexical access process is tied with the perceptual system gradual learning to recognize speech, in the sense that, as argued by (Mg and Wd., 1996b), that the process of learning to compensate for phonological variant forms (assimilated forms) depends on the gradual connection of the constraints involved in the goodness-of-fit between the variable output and input computation in lexical access(p.416). Second, they contend that unassimilated forms are more likely to be recognized by the recognition system as being the underlying forms and, thereby, inferring the assimilated forms (p.416). This is motivated by the fact that the phonological rules regulating assimilation of place are constrained by the phonological context viability for assimilation. For instance, the following phonological rules regulating six assimilatory patterns are central to the second assumption.

Based on the aforementioned IPM assumption, simplified network recognition (SNR) has been developed as figure (2) shows:

The SNR architecture has been generated to examine the lexical access process using a simple recurrent network trained on mapping a speech phoneme input, undergoing assimilation variation onto an output window, exploring a back propagation-algorithm. However, this research, as differently from Gaskells SNR hypothetical and

t → p / -# (p, b, m)
d → b / -# (p, b, m)
n → m / -# (p, b, m)
t → k / -# (k, g)
d → g / -# (k, g)
n → ŋ / -# (k, g)

Figure 1: Six Assimilatory Patterns (Gaskell et al., 1996)

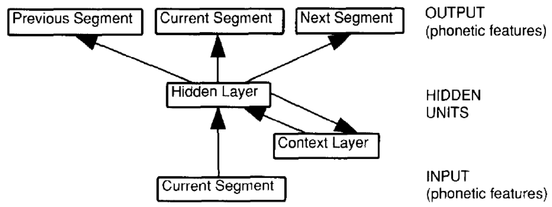

Figure 2: SNR Architecture (Gaskell et al., 1996)

practical orientations, tends to account for the assimilation processing by altering the current architecture drawing on the following perspective: the perceptual speech might have an active role in shaping regular and systematic phonological changes (e.g., assimilation of place process) and, therefore, would be able to shape the phonological rules (represented in figure 1) governing assimilation of place. Yet, the viability of the current perspective seems to be bound with the specific learning experience that the system undertakes; the system should be trained on mapping from underlying segments onto their surface assimilated segments. By doing so, the system is taken to learn the conditions in which the different six patterns of assimilations take place and, therefore, learn to recognize the respective phonological rules governing each pattern. This assumption is central to the current research revised SNR structure:

The proposed modified structure incites the research procedure described in the next section.

3 Research Procedure

The network will be trained on mapping from underlying unassimilated segments onto their surface assimilated variants in the prediction that the network would recognize the phonological rules underlying the assimilation of place pattern for the six cases as represented in figure (1). Moreover, following (Mg and Wd., 1996b) practice, in the purpose of stimulating the network to learn

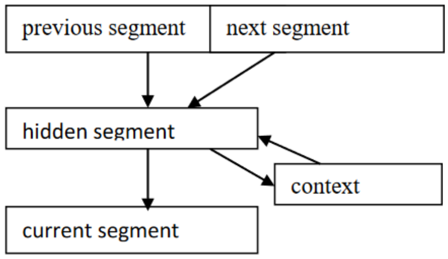

Figure 3: SNR Modified Structure

the associations between unassimilated segments (underlying segments) and their neighboring segments, we will incorporate a large corpus, Brown Corpus (H and N, 1960), whereby place assimilation will be artificially introduced into the corpus. A random selection of 50 % of coronal segments (/t/, /d/, /n/) contained in the data found in phonological contexts triggering assimilation (/p/, /b/, /m/, /k/, /g/) will be carried, then changing their place features so that to match them with the following phonological rules listed in figure (1). Moreover, the network will be trained on phonetic transcription of the artificially introduced assimilated words, as most of transcriptions of speech do not show phonetic representations of phonologically (assimilated) words. For this purpose, we will employ the LUND corpus of Stvartvik and Quirk (1980). We will convert the orthographic forms into phonetic forms employing a translation program (phonetizer program), and if any errors occur, they will be manually corrected. The network will be trained on 50 sweeps among the whole corpus, with the identity of the assimilated segments altering between the sweeps, employing TLearn program software.

4 Research Design

In the aim of examining the variant speech perception in terms of the phonological related-experience aspect, the process through which the perceptual system applies phonological rules to a range of underlying speech forms so that to recognize the surface forms, this research will assign an identification-based test to the network. Our predictions on the network performance pertain to the thought that it would make use of the recurrent links to apply the set of phonological rules regulating place assimilation to the data. For instance, the network is expected to shape the rule that for an underlying coronal (/t/) to be surfaced as (/p/), it must be followed by a segment with a bilabial non-coronal feature (/p /, /b/ or /m/). To test the network perceptual performance, we will use a set of two-word stimuli for 15 tokens and other 15 sentence-fillers to the trained network segment by segment, whereby the first word and second word final segments are deleted, however, the following words initial segments alter between triggering regressive assimilation of place to take place or blocking the process. For instance, the two-word stimuli /kwaipein / and /kwaisein/ derived from the word quite pain have their final coronal /t/ deleted ,however the /p /segment in the following word triggers the assimilation of the deleted segment as it provides a phonological context viable for the process to take place, while the second context, where /s/ is the initial consonant in the following word /sein/ blocks the assimilation of the deleted segment regarding the phonological context unviability for the process to occur. Each time, when the two-word stimuli for the 15 tokens are presented, they will be synchronized with the prime display (the tokens underlying forms). The network performance on the final segment in the first and second word will be measured at two points: one when the final deleted segment is recovered, whereby the output of the current segment window is recorded and examined, the other one will be measured when the following segments, presented along with the prime and the activations of the previous segment window were recorded so that to enable examination of the network identification of the final segment in the first word. Moreover, in the aim of testing the phonological context effect on the network perceptual performance, we will assign a discrimination-based test to the network with the use of other 15 carrier pair of words. However, this time the tokens will be presented with three modification types of the following phonological contexts: segments presented in their assimilated forms and, therefore, inserted in a minimally unviable phonological context (/kwaipein/), segments presented in maximally unviable phonological conditions, whereby the final segment in the first word is unassimilated and the initial segment in the following word generates non-lexical input (/kwaitkein/), and segments inserted in a viable phonological context for assimilation whereby the final segment is presented as unchanged and the initial segment in the neighboring word triggers the assimilation process (/kwaitpein/). Each type of these three variations will be presented twice to the network, followed by a mean coronal score across segments featured with velar or labial resultant place, to be used later in the analysis.

5 Conclusion

Given that assimilated-speech variants (speech sound-streams ending in the coronal/ t /, /d /, /n/ are neutralize respectively into /p/, b/,/m/,/k/ ,/g/ ,//, when followed by one of the non-coronals /p/, /b/, /m/ , /k/, /g/), this invokes a propagated activation throughout the interconnected input, output and hidden units of the system. In parallel with this, and in light of the SNR structure, which explores a back propagation-algorithm, the system tends to infer, through a gradual training on mapping from the surface (output) to the underlying (input) unit, the phonological rules underlying the assimilation process. Gaskell (Mg and Wd., 1996b) maintained, this simple architecture allows the network to learn generalizations based on the statistics of the speech stream, using these generalizations to improve the networks performance in the prediction of upcoming phoneme (p.415). However, the current research tends to improve the networks performance by resorting to a different mapping; training the network to map from input to output units, henceforth, optimizing the system to shape and accommodate the range of phonological variations to the process of the lexical access regularities.

Acknowledgments

The author would like to thank the almighty God for having a number of issues addressed in this paper.

References

Gaskell and Gareth. 2003. Modeling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*, 31 (3):447–463.

Kucera H and Francis N. 1960. Brown corpus. *US Cooperative Research Program Office of Education and Brown Corpus*, (p41-50).

ELMAN JEFFREY and McClelland JAMES. 1988. Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, 2.

Wilson WD Marslen. 1998. Mechanism of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 24(2).

James A MecQueen, Norris D, and Cutler A. 1999. Lexical influence in phonetic decision mismatches making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, 25(2).

Gaskell Mg and Marslen-Wilson Wd. 1996a. Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22(1):144–158.

Gaskell Mg and Marslen-Wilson Wd. 1996b. Phonological variation in speech perception: A connectionist approach. York, England.

Timothy T Rogers. 2009. Connectionist models. *Encyclopedia of Neuroscience*, (p75-82).

Matthew J Traxler. 2012. *Introduction to Psycholinguistic Understanding of Language Science*. Wiley-Black Well, India.