Perception of stress related vowels in Brazilian Portuguese

Leonardo Oliveira and **Eleonora Albano**

Phonetics and Psycholinguistics Laboratory (LAFAPE), C.P. 6045
State University of Campinas, Campinas, SP, 13010-050, Brazil
leocouto@iel.unicamp.br, albano@unicamp.br

Abstract

We investigated the influence of lexicality, stress, and vowel amplitude in the perception of vowel allophony in Brazilian Portuguese using an *ABX* speeded discrimination task. Stress position and lexicality seem to influence the detection of incorrect allophones. We discuss these findings in the context of possible extensions of Articulatory Phonology, a phonological model which can predict lexically specified allophones and gradient word properties.

1 Introduction

As a model of speech production, Articulatory Phonology ([1], henceforth AP) is particularly successful in describing phonetic detail in speech production and also in explaining gradient phonological alternation under two simple unifying principles: 1) different degrees of gestural overlap and; 2) different degrees of gestural magnitude.

However, these same principles are not sufficient, at least at the current state of the theory, to explain categorical allophonic variability, a phenomenon traditionally assigned to the phonological domain and possibly occurring in all languages. Since strict symbolic derivation is not a possible account for allophony within a dynamical framework, categorical alternation might be explained on the basis of positing different articulatory gestures for different allophones [2]. This possibility raises at least one major issue to be addressed by AP: the necessity of a phonetically detailed lexical representation which predicts categorical allophony and possibly other low level phonetic details. This fine-grained representation challenges a more traditional representation of abstract phonological units, like segments or features, which apply throughout the lexicon. Such a representation requires a more powerful lexicon to encode information like categorical allophony and, possibly, also word frequency, traditionally considered a matter of performance by Generative Phonology.

Evidence that low-level phonetic detail is linguistically specified in production and perception comes from different sources. Vowel-to-vowel coarticulation [3, 4], nasalization [5] and vowel quality [6] have different patterns across languages. Hay [7] provides evidence that frequency of occurrence of morphological structures influence /t/ deletion differently in words like *swiftly* and *softly* according to the frequency of its base and derived forms. Words that are more frequent than their bases tend to have more /t/ reduction than words that are less frequent than their bases.

For the particular case of categorical allophony, a study by Whalen and colleagues ([8], henceforth WBI) explored the [p $\mathrm{p^h}$] allophony in American English. The authors used discrimination, rating, and imitation tasks to assess the perception and production of [p $\mathrm{p^h}$] allophones in words and non-words in a series of five experiments. Subjects were asked to rate the degree of foreign accent of the stimuli, to discriminate them in an *AXB* paradigm, and to imitate them from the recording. These were distinct tasks, all using stimuli with /p/ allophones in correct and incorrect position in words and non-words. The stimuli used were produced by a trained phonetician in an attempt to control for naturalness. The results indicate that speakers can discriminate and imitate allophones in both correct and incorrect position within words and non-words. However, when subjects were asked to rate the degree of foreign accent, they gave higher ratings to words with allophones in correct position than in incorrect position. The same did not occur when phonotatically possible non-words were rated. This last result also poses a challenge to Generative Phonologies, which usually predict lexical allophony to be a phonological process that may even be morphologically conditioned, but in a categorical way.

More recently, a similar experiment by Jones [9], also for American English, compared naturalness ratings for aspiration in the [t] *vs.* [st] distinction and obtained results that conflict with WBI's. Jones' experiment did not show any difference for the word/non-word distinction in the preference for the correct position allophone, even including controlling factors like degree of familiarity for words and degree of naturalness for non-words. However, the presentation of the stimuli was bimodal, i.e., a previous orthographic form was presented before the auditory stimulus. Because there is evidence that orthographic forms activate phono-

ISBN 1-876346-48-5 © 2003 UAB

logical representations [10], the possible lexical effects could have been covered by this orthographic influence. Thus, Jones's results can hardly be compared to WBI's, since the former is not purely auditory.

Here, we further investigate the perception of [a ɐ] allophony in Brazilian Portuguese, presenting new data and discussing the key role of the lexicon in the influence of phonetic detail. We will also try to shed light on the allophony perception controversy.

2 Methods

2.1 Stimuli

Our corpus consists of words and non-words with [a ɐ] allophones either in correct position or with one allophone in an incorrect position. We chose the [a ɐ] allophony in Brazilian Portuguese because it has been carefully described from an acoustic point of view [11] and is conditioned mainly by stress position, similarly to aspirated *vs.* non-aspirated [p] in American English.

In order to obtain as natural as possible stimuli, we concatenated the vowel allophones in the incorrect position, controlling as far as possible some of its acoustic properties. We recorded every segment in its natural occurrence position in words and non-words spoken by a male speaker at a carefully controlled speech rate. The vowels were later concatenated in a base stimulus using the PSOLA algorithm of the *Praat* software to generate forms with allophones in incorrect position. The rationale for using a speech synthesis technique was to obtain stimuli that sounded natural and were produced in a less artificial situation than WBI's.

The word-like stimuli consisted of two forms of the verb *pagar* (to pay), a high frequency verb with contrastive stress in the last syllable: [pa'ga] (the infinitive form) and ['pagɐ] (a present indicative form). The non-words forms, also with contrasting stress, are ['vapɐ] and [va'pa]. Besides, we had variants of both words and non-words with allophones in incorrect position. The forms and variants are (the underlined vowel is the incorrect):

⇒Word forms:
['pagɐ], ['pag<u>a</u>], ['p<u>ɐ</u>gɐ];
[pa'ga][1], [pa'g<u>ɐ</u>], [p<u>ɐ</u>'ga].
⇒Non-word forms:
['vapɐ], ['vap<u>a</u>], ['v<u>ɐ</u>pɐ];
[va'pa], [va'p<u>ɐ</u>], [v<u>ɐ</u>'pa].

The fundamental frequency and vowel duration were kept controlled in all forms. A previous experiment suggested that different conditions of vowel duration did not affect the stimuli discrimination [12]. The stimuli pairs were matched for phonological neighborhood density and phonotactics. The neighborhood density was estimated by calculating for each form the sum of the $\log_{10}$ frequencies of each phonological neighbor:

$$\sum_{j=1}^{n}\left(\log_{10}\left(\mathrm{Freq}_{N_j}\right)\right) \quad (1)$$

As a rough measure of phonotactics, we calculated the $\log_{10}$ of frequency-weighted diphones:

$$\log_{10}\left(\sum_{j=1}^{n}\left(\mathrm{Diphone}_j * \mathrm{Freq}_{diphone_j}\right)\right) \quad (2)$$

2.2 Experimental design

In order to investigate the possible lexical effects on allophony perception, we used a speeded *ABX* discrimination task with reaction time measurements. Subjects had to respond as fast as possible whether the third word in the series was identical to the fist or the second. The reason for using discrimination responses and reaction time measurements was to avoid possible post-lexical effects and obtain more objective measures. We avoided ratings by degree of foreign accent because we were not sure whether this measure would be reliable in Portuguese. The *ABX* discrimination was preferred over the *AXB* because it was not possible to have stimuli with the same durations and to force a single response strategy. Even recording the reaction time after the second stimulus in the triplet it was possible that subjects responded only after the third stimulus. With the *ABX* paradigm, we tried to guarantee a more uniform decision strategy, since the subject needs to listen to the third stimulus to make a decision.

The controlled experimental conditions were: 1) Three 2×2 combinations of stimuli variants; 2) Word/non-word distinction; 3) Two different stress conditions, which are: a) penultimate canonical stress; and b) final stress; and 4) two amplitude patterns: a) vowels with same amplitude in both stress positions; and b) vowels with canonical amplitude differences between stress positions.

The dependent variables were: 1) Correct/incorrect answers; and 2) reaction time.

2.3 Experimental procedure

For the presentation of the stimuli we had a set of three word and three non-word forms, each in two different stress conditions, plus two amplitude patterns, and eight repetitions for each triplet, totaling 192 trials presented in 8 separate blocks. The third stimulus in the triplet was the same as the first in half of trials and the same as the second in the other half. Each stress condition and amplitude pattern was presented separately. All subjects responded to all blocks, the block order presentation being random between subjects. A total of 18 subjects participated in the experiment.

[1]The infinitive form is also pronounced [pa'gaɹ] in careful speech.

ISBN 1-876346-48-5 © 2003 UAB

Each stimulus in the *ABX* group was presented with a 500 ms interval from each item and the triplets were presented with a 2 second interval after the subject response. There was a time limit of two seconds for each response. The subjects' correct and incorrect responses and their reaction times were measured for each trial. For the presentation of the stimuli we used the DMDX [13] software.

3 Results

3.1 Correct/Incorrect responses

A Chi-square test showed a significant overall effect for the correct/incorrect responses patterns for words and non-words ($\chi^2(1) = 7.99, p = 0.0047$). Subjects made more errors responding the non-word triplets than words.

The stress position factor also showed an effect in the number of incorrect responses ($\chi^2(1) = 48.28, p < 0.001$). Subjects made more errors responding non-canonical last syllable stress than penultimate stress.

The two amplitude conditions did not have a different pattern for the correct and incorrect responses.

3.2 Reaction time responses

The analysis of reaction time data included only the correct response times. The inclusion of incorrect responses would introduce an additional source of variability in the data, which already presents a right skewed distribution [14] (Kurtosis = 4.97 and Skewness = 1.25). Moreover, the incorrect response times come from different sources such as real misidentifications, distractions, and errors in button pressing.

Before the analysis, the reaction time data were normalized, using the Box-Cox transformation:

$$y = \frac{x^\lambda - 1}{\lambda} \tag{3}$$

Where x is the value to be transformed and λ is an empirically determined parameter. For our data $\lambda = 0.1932$. The transformed data has Kurtosis = 3.5 and Skewness = -0.00007. The reason to apply a nonlinear transformation to the data was to be able to meet the assumptions of normality of parametric statistics. The raw data were also analyzed with the same statistical tests and the significant differences remained. We will report only the results for the normalized data.

We performed a repeated measures ANOVA, according to the recommendations of [15], considering each subject as the a single case. Due to missing data resulting from timeouts and incorrect responses frequently found in some of the eight trials of each triplet, we decided to include five randomly selected valid trials. In about ten percent of cases, we had to include the mean reaction time of each subject to complete the five valid trials. This was necessary because we needed to have all cells with valid data to perform the analysis. To check the influence of this procedure, we performed the analysis with the mean reaction time of the eight trials and obtained the same results. We will just report the results for the data with five trials. This way, we had as repeated measures main effects the 5 levels of trials $\times$ the 3 levels of combination of stimuli $\times$ 2 levels of lexicality $\times$ 2 levels of stress position $\times$ 2 levels of amplitude patterns.

The repeated measures ANOVA showed an effect only for the stress factor ($\text{F}(1, 17) = 12.39, p = 0.002$), sphericity not being an issue, since the factor had only two levels. Subjects discriminated canonical penultimate syllable stress faster than last syllable stress.

4 Discussion

The discrimination task results (i.e. incorrect allophones are harder to discriminate in non-words) and the reaction time results (i.e. words and non-words do not have different reaction times) are somewhat ambiguous. Lexicality (and in this case word frequency, since the verb form is a high frequency word and the non-word can be considered a very low frequency lexical item) does have an influence in the response accuracy. But when the correct response times are analyzed separately, the differences disappear. The incorrect response pattern is predictable considering psycholinguistic studies of word frequency: lexicality and frequency seem to favor, at a level above chance, the discrimination of forms with incorrect allophones. The absence of the same effect in the reaction time for the correct responses can indicate a difference in the processing of the incorrect forms that deserves further inquiry.

The correct/incorrect response results are compatible with WBI's (who found differences in rating depending on lexical status) and contrast with Jones's, who found no differences in rating depending on lexical status. Our results support the postulation of categorical allophones in lexical entries, since different allophones seem to be discriminated at a level higher than chance depending on lexical status and possibly word frequency.

A phonological model which incorporates low-level phonetic detail is proposed by Pierrehumbert [16, 17]. Pierrehumbert's proposal is grounded on exemplar theories of perception extended to speech perception and production. In exemplar theory, categories are represented in long term memory by accumulation of tokens for each category, which can vary along the relevant physical dimensions of speech. This way, factors like frequency of occurrence and probabilistic phono-

ISBN 1-876346-48-5 © 2003 UAB

tactics are automatically predicted by the model in a non-categorical fashion. The lexicon is no longer a separate module, but exhibits gradient characteristics which affect the production and perception of speech. Our results, particularly the lower reaction times and higher discrimination accuracy for canonical (and more frequent) penultimate stress, also seems to be compatible with Pierrehumbert's proposal, since words with canonical stress should exhibit more tokens, facilitating lexical access in canonical stress position. AP key features and Pierrehumbert's proposal are compatible [2, 18] in the sense that both models can handle linguistically meaningful phonetic detail and incorporate the pervasive effects in of word frequency, probabilistic phonotactics and lexical neighborhoods speech production and perception largely described in the literature.

Further evidence of the influence of lexical status can be brought to bear on this issue by testing a larger number of words as to how frequency, lexical neighborhood density and probabilistic phonotactics influence the perception of allophony.

Acknowledgments

Work supported by a master's degree scholarship grant from FAPESP (01/00137-9) in connection to the project "Integrating continuity and discreteness in modelling phonic and lexical knowledge", funded by FAPESP (01/00136-2). Thanks to Edson Françozo, Plínio Barbosa, Ana Luiza Navas, Fábio Violaro, and Beatriz R. de Medeiros for helpful suggestions. Special thanks to our speaker and the subjects who participated in the experiment.

REFERENCES

[1] C. P. Browman and L. Goldstein, "Articulatory phonology: an overview," *Phonetica*, vol. 49, pp. 155–180, 1992.

[2] E. C. Albano, *O gesto e suas bordas: esboço de Fonologia acústico-articulatória do português brasileiro*, Mercado de Letras, Campinas, 2001.

[3] S. E. G. Öhman, "Coarticulation in VCV utterances: spectrographic measurements," *Journal of the Acoustical Society of America*, vol. 39, pp. 151–68, 1966.

[4] S. Y. Manuel, "The role of contrast in limiting vowel-to-vowel coarticulation in different languages," *Journal of the Acoustical Society of America*, vol. 88, pp. 1286–1298, 1990.

[5] H. Clumeck, "Patterns of soft palate movements in six languages," *Journal of Phonetics*, vol. 4, pp. 337–351, 1976.

[6] S. F. Disner, *Vowel quality: the relation between universal and language specific factors*, Ph.D. thesis, University of California at Los Angeles, Los Angeles, CA, 1983.

[7] J. B. Hay, *Causes and consequences of word structure*, Ph.D. thesis, Northwestern University, Evanston, IL, 2000.

[8] D. H. Whalen, C. T. Best, and J. Irwin, "Lexical effects in the perception and production of American English /p/ allophones," *Journal of Phonetics*, vol. 25, pp. 501–528, 1997.

[9] A. K. Jones, "A lexicon-independent phonological well-formedness efeect: listener's sensitivity to inappropriate aspiration in initial /st/ clusters," *UCLA Working Papers in Phonetics*, vol. 100, pp. 33–71, 2002.

[10] M. T. Turvey and G. Lukatela, "An evaluation of the two-cycles model of phonology assembly," *Jounal of Memory and Language*, vol. 42, pp. 183–207, 1999.

[11] P. A. Aquino, "O papel das vogais reduzidas póstônicas na construção de um sistema de síntese concatenativa para o português do brasil," M.S. thesis, Instituto de Estudos da Linguagem — Universidade Estadual de Campinas, Campinas, 1997.

[12] L. Oliveira, "Perception of /a/ allophony in brazilian portuguese," june 2002, Poster presented at the Eighth Conference on Laboratory Phonology.

[13] J. C. Forster and K. Forster, "DMDX version 3," 2001.

[14] R. H. S. Carpenter and M. L. L. Williams, "Neural computation of log likelihood in the control of saccadic eye movements," *Nature*, vol. 377, pp. 59–62, 1995.

[15] L. Max and P. Onghena, "Some issues in the statistical analysis of completely randomized and repeated measures designs for speech," *Journal of Speech, Language, and Hearing Reasearch*, vol. 42, pp. 261–270, 1999.

[16] J. Pierrehumbert, "Word-specific phonetics," in *Laboratory Phonology VII*, C. Gussenhoven and N. Warner, Eds., pp. 137–157. Mouton de Gruyter, Berlin, 2002.

[17] J. Pierrehumbert, "Exemplar dynamics: Word frequency, lenition, and contrast," in *Frequency and the Emergence of Language Structure*, J. L. Bybee and P. Hopper, Eds., pp. 137–157. John Benjamins, Amsterdam, 2001.

[18] A. Kochetov, "Syllable position effects and gestural organization: articulatory evidence from russian," Paper presented at the 8th Conference on Laboratory Phonology, 2002.