ACOUSTICS vs. PHONEMES IN LEXICAL ACCESS

William Barry and Bistra Andreeva

Institut für Phonetik, Universität des Saarlandes
wbarry/andreeva@coli.uni-saarland.de

ABSTRACT

Phonetic perception and lexical access is sensitive to acoustic traces of co-articulatory processes in overlapping neighbouring segments. Longer distance coarticulatory effects, though well documented in production studies, have not been examined with regard to their contribution to lexical access. Using an eye-tracking paradigm, we examines whether the acoustic reflex of anticipatory lip-rounding and lip-spreading in initial /ʃ/ in German CC and CCC word onset clusters is used to decide between lexical candidates with phonemically identical onsets prior to the contrasting vowel. The results show a clear effect of the pre-vocalic consonant information, but the effect is not symmetrical for rounded and unrounded /ʃ/. Results are discussed in relation to a phonemic vs. a (demi-)syllabic basis of lexical decisions and markedness theory.

Keywords: Eyetracking, lexical access, phoneme, syllable, co-articulation.

1. INTRODUCTION

Speech perception is a complex operation that humans perform with relative ease in widely differing situations, supported by background knowledge, contextual information and visual cues which strongly constrain the meaning. These factors make understanding even under adverse acoustic conditions quite robust. So, precisely defined acoustic properties, and the speech-sound categorization which these could support, are clearly not *indispensible* to word recognition. Word meaning is such a dominant aspect of perception that in sound categorisation tasks a meaningful unit at one end of an acoustic continuum between two phonemic categories shifts the boundary between the two categories to the disadvantage of the non-meaningful unit [5] and truncation of a continuum between two meaningful units at a point within the ambiguous boundary values does not only *not* prevent acceptance of the ambiguous stimulus as a word; it also causes a temporary shift of the category boundary in a subsequent categorisation task using the whole continuum [7].

But perception tasks without the support of shared knowledge, visual and contextual cues have shown beyond doubt that we *do* make use of fine acoustic detail in the speech signal to recognize the meaning of words. There is ample experimental evidence from sub-phonetic mismatch experiments that we use the fine effects of co-articulation in the ongoing acoustic input both to categorize sounds and to predict up-coming parts of a word [8, 10, 11, 9, 2].

A widespread assumption in psycholinguistics, congruent with these observations (see [6] for a discussion), is that the incoming signal activates (all) the potential word candidates that are not ruled out by the current acoustic structure, and the number of candidates dwindle as additional input arrives which no longer supports part of the original cohort. Co-articulatory information speeds up this selection process by allowing information about structurally sequential sub-units of the word to be processed in parallel (e.g. aspiration after the release of a plosive contains information on its [-voice] status, its place of articulation *and* the quality of the following vowel).

An unresolved issue linked to the multiple activation assumption is whether the exclusion of non-matching candidates requires the phonemic categorization of the incoming signals. Co-articulatory mismatch evidence is often used as an argument *against* the phoneme as the unit operating in this process, but as Cutler [1] points out, the evidence is neutral. For example, within an Action Theory framework [4], i.e., under the assumption of sequenced but overlapping phonemes, evidence of a contribution from *neighbouring* phonemes is totally compatible both with a phonemic categoriz-

ation theory and results from co-articulatory mismatch experiments.

This study examines the contribution of the sequential phonemic structure to the process by considering the effect of *more distant* co-articulatory information on lexical access. Rounded and unrounded allophones of initial /ʃ/ in two- and three-consonant clusters are tested for their contribution to the early recognition of words which are phonemically identical up to the rounded or unrounded vowel following the cluster. It is well known that sibilants vary in their spectral energy distribution as a function of the following vowel. Whalen [10] demonstrated that the vowel category can shift the /s – ʃ/ category boundary in sibilant + vowel sequences. Lexical access has also been shown to be disturbed by incorrect /s – ʃ/ quality in both vowel + sibilant and sibilant + vowel sequences [11].

Words like *Stiefel* (/ʃtiːfəl/, Engl. *boot*) and *Stuhl* (/ʃtuːl/, Engl. *chair*) exhibit anticipatory colouring of the initial /ʃ/ which can potentially signal the upcoming /iː/ or /uː/. In a word recognition process that proceeds with the categorization of phonemic units, the allophonic /ʃ/-colouring has no function. The recognition of *Stuhl* vs. *Stiefel* cannot be decided before the onset of the vowel. In perceptual models which are not constrained by phonemic categorization, the quality of the /ʃ/ would be predicted to aid word recognition prior to the vowel onset.

2. METHOD, MATERIAL AND ANALYSIS

The time course of word-recognition was investigated using the eye-tracking paradigm. This method captures the gaze fixation reflex from the word recognition process from the moment the acoustic input starts. It also has the advantage of avoiding any meta-level description (whether phonetic or orthographic). In addition, it is not sensitive to (experimentally irrelevant) differences between competing recognition candidates that come later in the word (due to word length), an important factor given the dearth of suitable words. The presentation of a tableau with four pictures (a target, a competitor and two phonetically unrelated distracters) primes four words, thus weakening all the members of the cohort activated by the acoustic input except the target and the competitor.

2.1. Material

Twenty pairs of testwords were prepared for presentation. The pairs consisted of picturable items with one word beginning with /ʃ/+C(C)+ *rounded* vowel and the other beginning with the same /ʃ/+C(C) sequence + an *unrounded* vowel.

To present the testwords, there were 12 tableaux with /ʃ/+C-onset target and competitors (*Stuhl*, *Stiefel* etc.), 8 with /ʃ/+ CC-onset targets and competitors (*Strumpfhose*, Engl. *tights*; *Straßenbahn,* Engl. *tram* etc.). A further 12 with vowel-onset targets (*Uhr*, Engl. *clock/watch*; *Igel*, Engl. *hedgehog* etc. were used to provide a word recognition baseline, and there were 12 distracter targets (not evaluated) which, together with the vowel-onset targets, provided a balance for the words with /ʃ/ onset. Three unevaluated training items were given at the beginning.

The acoustically presented instruction to the subjects was: "*Klick' auf das folgende Bild* __" (Engl.: "*Click on the following picture* __" cf. Soundfiles/1.wav and 2.wav), with the name of the pictured object after a 500 ms pause. The instruction sentence and the picture names were spoken by the same person.

The acoustic signal of the picture names was controlled in the following way: One clearly rounded and one unrounded [ʃ]-segment were selected for the alveolar and for the bilabial context (/ʃt_/ and /ʃp_/). For all stimuli, the duration of the [ʃ] segments were equalized to 150 ms, giving an average onset duration of 230 ms for the ʃCV_ and 260 ms for the ʃCCV_ stimuli. The amplitude of the [ʃ] segments was adjusted so that 5 phonetically trained listeners judged them to be equally loud when presented in isolation.

2.2. Subjects and Procedure

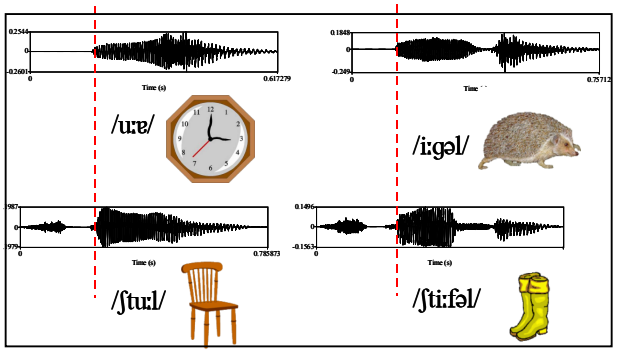
In total 44 subjects were tested, 12 in a pilot study to test the feasibility of the task and 32 in the main experiment. Both sets of subjects were sub-divided such that all the groups were presented with half the targets containing rounded vowels (with unrounded-vowel competitors) and the other half containing targets with unrounded vowels (and rounded-vowel competitors). For two groups, a particular tableau was used for the rounded-vowel target and for the other two groups, it was used for the unrounded-vowel targets with the position of

target and competitor swapped (see accompanying file Graph.pdf 1 and compare soundfiles/1.wav and 2.wav).

After calibration of the SMI head-mounted eyetracker with Eyelinx software, each subject was presented with the 47 tableaux. Their instructions were to focus on the cross situated symmetrically between the pictures located in the four corners of the screen until told to click on one of the pictures. Fixation points were recorded every 10 ms from 200 ms prior to vowel onset for all targets (this was on average 30 or 60 ms after frication onset, depending on the ʃCV_ or ʃCCV_ structure of the target word). The target, competitor or one of the distracters were registered as fixated if the gaze fell within a quadilateral area surrounding each of the pictured objects.

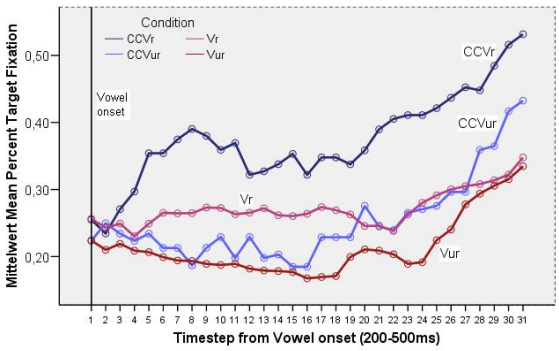
The null-hypothesis is that the differing acoustic reflex in the initial /ʃ/ will not affect word recognition speed. Therefore preferential fixation on the target will occur at the same time relative to vowel onset, whether the word begins with ʃCCV_, ʃCV_ or with V_. The reference point for comparing fixation curves will therefore be the onset of the vowel (see fig. 1 and cf. Soundfiles/3.-6.wav). Since intentional gaze fixation takes approx. 200 ms to carry out, reactions to information contained in the first 50 ms of the vowel can be expected about 250 ms from vowel onset. Target fixation significantly prior to this point must be attributed to information processed prior to vowel onset

Figure 1 Reference points for comparing fixation curves in words beginning with a vowel or a consonant cluster



Evaluation of the fixation curves was based a) on the point at which the target-fixation curve diverge from the competitor fixation curve, b) on a one-way ANOVA with stimulus type as independent variable and the target-fixation scores for each 10 ms step during the window 100-300 ms after vowel onset as dependent variable. The latter procedure was based on the rationale that the fixation curve rises earlier for an earlier recognition point and the resulting average fixation score for the critical window is higher than for a later recognition point with a consequent later rise in the fixation curve (see figure 2).

Figure 2 Fixation curves for ʃCVur/r (unrounded and rounded) and Vur/r targets



3. RESULTS

The point of divergence in the fixation curves of the target and the competitor for words with vocalic onset was 245 ms after vowel onset for words beginning with an unrounded vowel and 230 ms for those with a rounded vowel (see Graph.pdf, figs. 2 and 3). These values are closely comparable to those observed in the pilot study (270 and 240 ms, respectively.

For ʃCV_ stimuli the point of divergence was 235 ms for words with *un*rounded vowels and 15 ms for words with *rounded* vowels (though the difference between target and competitor fixations remains relatively small for another 170 ms, with another strongly divergent movement at 185 ms; see Graph.pdf, figs. 4 and 5). The corresponding values in the pilot study were 240 and 170 ms.

For ʃCCV_ stimuli the point of divergence was 250 ms for words with *un*rounded vowels and 95 ms for words with *rounded* vowels (though the difference between target and competitor fixations again remains relatively small for another 75 ms, with the final strong divergence at 170 ms). The values in the pilot study were 280 and 200 ms.

These results show a clear effect of the rounded /ʃ/ but little effect of the unrounded /ʃ/. Table 1, which shows the stimulus groups with significantly different fixation curves, provides statistical support for this finding.

Table 1 Post-hoc separation of stimulus types after one-way ANOVA based on proportional fixation score

Significantly different sub-groups			
Stimulus type	1	2	3
V-unrounded	0.219		
CCV-unrounded	0.267	0.267	
V-rounded		0.278	
CCCV-unrounded			0.389
CCV-rounded			0.396
CCCV-rounded			0.402

The ʃCV stimuli with unrounded vowels do not differ from the vowel-onset stimuli, for which (in the nature of the stimulus structure) no acoustic information is received prior to vowel onset. The ʃCCV stimuli with unrounded vowels group with the two stimulus types with rounded vowels

These results differ only slightly from those of the pilot experiment (see Graph.pdf, no. 8), where the ʃCV and ʃCCV stimuli with unrounded vowels group together and are both significantly different from both the vowel-onset stimuli and the ʃC(C)V stimuli with rounded vowels.

4. DISCUSSION

The quicker rise times of the target fixation curve for stimuli with precursor consonant clusters than for vowel-onset stimuli clearly indicate that the information in the initial /ʃ/ relevant to the disambiguating vowel is being used to access the lexicon. Since the /ʃ/ is separated from the vowel by one or two other consonants, the contribution cannot be seen as a part of a sequential phonemic categorization process. Whether or not the /ʃ/ is recognized as a separate phonemic unit, the subphonemic frication colouring associated with the later vowel is already helping to disambiguate the syllabic unit (or at least the demi-syllabic unit (Dupoux 1993).

The stronger effect of the co-articulatorily *rounded* /ʃ/ frication indicates a greater perceptual salience of [+rounded] compared to [−rounded]. For this there is no *a priori* psychoacoustic reason. In fact, from the known default tendency of /ʃ/ to be produced with slightly rounded lips, it would be more plausible to expect the *unrounded* allophone to be more strongly predictive of the later vowel quality. The observed asymmetry offers interesting empirical support for the perceptual validity of phonological markedness.

It must, however, be borne in mind that, although the eye-tracking data is a direct reflection of the semantic decoding process, it is in its very nature a "noisy" signal, particularly at the level of temporal resolution involved here. The variance behind the average values observed confound individual processing variance with inter-individual differences in "phonetic sensitivity" and reaction times. On the other hand, the largely congruent data obtained in the pilot and the main experiment suggest a robust effect. Additional evidence from other experimental paradigms (e.g., reaction times for both word and vowel recognition) is clearly also needed.

Acknowledgments: Grateful thanks to Andrea Weber for unstinting help and advice, and to Matt Crocker for the provision of technical support

5. REFERENCES.

[1] Cutler, A., Broersma, M. 2005. Phonetic precision in listening. In W. Hardcastle, & J. Beck (Eds.), *A figure of Speech: a festschrift for John Laver* (pp. 63-91). Mahwah, NJ: Erlbaum.

[2] Dahan, D., Magnuson, J., Tanenhaus, M. K., Hogan, E. 2001. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16*, 507-534.

[3] Dupoux, E. 1993. The time course of prelexical processing: The syllable hypothesis revisited. In: Altmann, G. Shillcock, G. *Cognitive Models of Speech Proce*ssing. 83-114. Hove: Lawrence Erlbaum.

[4] Fowler, C. A. (2003). Speech production and perception. In A. Healy and R. Proctor (eds.). *Handbook of psychology, Vol. 4: Experimental Psychology*. (pp. 237-266) New York: John Wiley & Sons

[5] Ganong, W.F., 1980. Phonetic categorization in auditory word perception, *J. Exp.Psych.:HPP*, 6:110-125.

[6] McQueen, J.M. 2003. Speech perception. In K. Lamberts & R. Goldstone (Eds.), *The handbook of cognition*. London: Sage Publications.

[7] Norris, D., McQueen, J.M., & Cutler, A. 2003. Perceptual learning in speech. *Cognitive Psychology*.

[8] Streeter, L.A., & Nigro, G.N. 1979. The role of medial consonant transitions i n word perception. *Journal of the Acoustical Society of America, 65*, 1533 -1541.

[9] Utman, J.A., Blumstein, S.E., & Burton, M.W. 2000. Effects of subphonetic and syllable structure variation on word recognition. *Perception & Psychophysics, 62*, 1297-1311.

[10] Whalen, D.H. 1984. Subcategorical mismatches slow phonetic judgments. *Perception & Psychophysics, 35*, 49-64.

[11] Whalen, D.H. 1991. Subcategorical phonetic mismatches and lexical access. *Perception & Psychophysics, 50*, 351-360.