Öhman, Sven. 1979. 'Introduction: Gunnar Fant, pioneer student of human speech'. In S. Öhman & B. Lindblom (eds), *Frontiers of speech communication research*, xv-xxv. London: Academic Press.

Roca, Iggy & Wyn Johnson. 1999. *A course in phonology*. Oxford: Blackwell.

Tønnessen, Finn Egil. 2002. 'ICT and communication (problems)'. In S. Strömqvist (ed.), *The diversity of languages and language learning*, 91-101. Lund: Lund University.

Per Henning Uppstad <Per_Henning.Uppstad@ling.lu.se>

Lund University, Dept. of Linguistics
Working Papers 51 (2005), 225–238

Listeners' sensitivity to consonant variation within words

Joost van de Weijer

1 Introduction

Part of our native language competence is the implicit knowledge of phonological word structure. Speakers of English know that *flink* is a possible word of English, but that *lfink* is not, because phonotactic constraints do not permit the combination *lf* as a word onset.

Experimental work with infants shows that this knowledge develops at a very early age. Jusczyk et al. 1993, for instance, demonstrated that infants at the age of nine months have knowledge of the sounds that occur in their native language. In this experiment, a group of American infants and a group of Dutch infants were tested using the preferential looking paradigm. The infants in both groups listened to Dutch and English words. The words were matched in terms of word length and stress pattern, but the Dutch words contained speech sounds that are not part of the English sound system, whereas the English words contained speech sounds that are not part of the Dutch sound system. At nine months of age, the American infants had a listening preference for the English words, and the Dutch infants had a listening preference for the Dutch words. At six months of age, there was no difference between the two groups, suggesting that sensitivity to the sound system of the native language develops between six and nine months of age.

In another study, it was shown that infants in the same period develop knowledge of phonotactic patterns in their native language (Jusczyk, Luce & Charles-Luce 1994). The infants in this study listened to nonsense words that contained combinations of speech sounds that were either highly probable or highly unlikely in their native language. The results of this study were similar to those of Jusczyk et al. 1993. Nine-month-old infants preferred to listen to stimuli with high-probable sound combinations but six-month-old infants had no preference yet.

Jusczyk, Cutler and Redanz 1993 tested infants' sensitivity to prosodic word structure. Previously, it had been demonstrated (Cutler & Carter 1988) that the predominant stress pattern in English is trochaic (strong-weak). Most English disyllabic words have this stress pattern, and these words are also more frequent than words that have stress on the second syllable. Jusczyk, Cutler and Redanz examined whether infants are sensitive to the distribution of stress pattern. They presented disyllabic words with stress on the first or the second syllable. Once again, infants at nine months of age had a listening preference for the words with the predominant stress pattern, whereas infants at six months of age did not yet have such a preference.

One of the questions related to the knowledge of phonological word structure is how this knowledge is related to the process of spoken word recognition. The mental lexicon consists of tens of thousands of entries and speech is often produced at a high rate. Nevertheless, comprehension of spoken language under normal circumstances does not lead to any apparent problems for the listener. And yet, the speech signal is characterized by a number of features that makes it extremely difficult for a computer to achieve what the human listener is able to so effortlessly.

One of the difficulties with the recognition of words that has received considerable attention is the problem of lexical segmentation. This problem is caused by the fact that there are no reliable cues to word boundaries in the acoustic signal, analogous to the spaces between the words in written language (Cole & Jakimik 1988). The question is how listeners identify the word boundaries in spoken language.

The solution proposed by Cole and Jakimik for the lexical segmentation problem is that listeners make use of their lexical knowledge for the identification of word boundaries. According to this view, lexical segmentation proceeds in a more or less left-to-right manner. Furthermore, the listener uses his or her knowledge about the vocabulary and the grammar of the language for identifying a word. Recognition leads to the identification of the onset of the next word.

Although this is an intuitively plausible solution, there are some problems that remain unresolved. First of all, as was pointed out by McQueen, Norris & Cutler 1994, the lexicon is characterized by a relatively high degree of overlap. Many short words occur as substrings within longer words (e.g. *in*, *wind*, *dough*, or *win* in *window*), so that these shorter words cannot be recognized with certainty until after their offset. A second question is: if word segmentation is triggered by word knowledge, how do prelingual infants learn where in an utterance the word boundaries are? Clearly, prelingual infants cannot rely on their knowledge of word meaning, or grammar for segmentation. Nevertheless, infants as young as seven and half months old are capable of recognizing a word when that word is presented in context, as was shown by Newsome & Jusczyk 1995.

Given these unresolved issues, it has been proposed that listeners follow explicit strategies for locating word boundaries in fluent speech. These strategies are based on non-lexical, language-specific regularities – such as prosodic word structure or phonotactic constraints on word form – which tend to correlate with the presence of word boundaries.

Evidence for explicit segmentation strategies comes from experiments using the word-spotting paradigm. In these experiments listeners have to detect target real words (e.g. *plan*) embedded in nonsense words (e.g. *plancil*). Using the word-spotting paradigm, Cutler & Norris 1988 found that listeners detected monosyllabic target words (e.g. *mint*) embedded in disyllabic weak-strong words (e.g. *mintev*) significantly faster than target words embedded in the beginning of disyllabic strong-strong words (*mintayf*). The result suggests that listeners make an attempt at lexical access every time they hear a strong syllable, which slowed down the process of detecting the target word in the experiment. It should be noted that this segmentation strategy, which apparently works well for English, does not necessarily work for other languages.

Similarly, McQueen 1998 investigated whether listeners use phonotactic information for segmentation. His experiment was carried out with Dutch materials and Dutch listeners. The target words were either aligned with the syllable boundary (e.g. *rok* 'skirt' in *fiem-rok*), or misaligned (e.g. *rok* in *fie-drok*). The alignment was caused by the phonotactic constraints of Dutch. The aligned target words were detected significantly faster than the misaligned targets, suggesting that the knowledge of phonotactic structure plays a role in the recognition of word boundaries. Also here, it should be noted that phonotactic regularities are language-specific.

The focus of the present study is a different aspect of phonological word structure. In a recent study (van de Weijer 2003), the hypothesis was tested that one aspect of 'word wellformedness' implies that the consonants that a simple, monomorphemic word is constructed of are different from each other. According to this hypothesis, for instance, the nonsense words *tandle* or *bandle* are better word candidates than *nandle* or *landle*.

Note that this hypothesis is not the same as the obligatory contour principle (OCP), a well-known phonological principle according to which adjacent identical segments are prohibited (Clements & Hume 1995). Initially, the OCP was evoked to explain suprasegmental patterns, but later also to segments. According to the hypothesis that is the focus of the present study, even consonants that are separated by more than a single vowel (as the two *l*s in *landle*) affect the wellformedness of a word.

In order to find support for this hypothesis, van de Weijer 2003 carried out a corpus study of Swedish. The corpus that was analyzed consisted of 5,388 monosyllabic and disyllabic word types which had a total token frequency of roughly 28 million tokens. The results showed that very few (only 1.57%) word tokens in this corpus contained two or more IC, demonstrating that it is indeed an unusual pattern.

The question addressed in the present study is whether listeners are sensitive to this aspect of word wellformedness. If listeners somehow respond differently to words with IC than to words without IC, then there is additional evidence that it is a true characteristic of words to be constructed of different consonants.

For this purpose, an auditory lexical decision task was carried out. In this type of experiment, the listener's task is to decide as rapidly as possible whether an aurally presented item is a real word or a nonsense word by pressing one of two buttons on a response box. The forms of the real words and the nonsense words were systematically varied. Some of the real words and the nonsense words contained IC, but others did not. The expectation was that, if listeners are sensitive to this aspect of word structure, the presence of two IC within a stimulus should speed up the rejection of that stimulus if it is a nonsense word, and slow down the recognition of that stimulus if it was a real word. The experiment was carried out in Sweden with Swedish materials and Swedish subjects.

2 Method

2.1 Stimuli – Real words

A total of 30 monosyllabic and 30 disyllabic real words were used as test items (see Table 1 for a complete list of the items). They were all common monomorphemic content words. In order to prevent frequency effects, the words were selected from a small proportion of a Swedish word frequency list. This list, provided by the University of Gothenburg, consisted of approximately 100,000 word types that were most frequent in a large corpus of written language. The test items used for the present experiment were all ranked lower than the top 1.2% of all the word types, and higher than the least frequent 90.0% of all the word types.

Table 1. Test items

real words		nonsense words	
monosyllabic	disyllabic	monosyllabic	disyllabic
bomb 'bomb'	*mörker* 'darkness'	*nand*	*nindel*
malm 'ore'	*koka* 'cook'	*nen*	*kunker*
sats 'clause'	*papper* 'paper'	*glagg*	*saster*
stolt 'proud'	*raster* 'screen'	*grög*	*dedel*
tät 'dense'	*skrika* 'cry out'	*dred*	*doder*
tält 'tent'	*korrekt* 'correct'	*tolt*	*vavel*
bror 'brother'	*tendens* 'tendency'	*tift*	*reger*
slips 'tie'	*lokal* 'room'	*mämt*	*palopp*
stat 'state'	*klocka* 'clock'	*tust*	*lanel*
sås 'sauce'	*skakel* 'shaft'	*pramp*	*fefir*
bank 'bank'	*handel* 'trade'	*därg*	*nasker*
brand 'fire'	*humor* 'humour'	*flad*	*säver*
damm 'dust'	*dotter* 'daughter'	*sord*	*bistel*
falsk 'false'	*lager* 'stock'	*forg*	*lonus*
grepp 'grasp'	*cykel* 'bike'	*grist*	*kuffel*
höjd 'height'	*focus* 'focus'	*kröst*	*rygel*
klang 'sound'	*läge* 'situation'	*garm*	*pensor*
makt 'power'	*monster* 'monster'	*kans*	*madel*
mynt 'coin'	*gratis* 'free'	*nikt*	*sotor*
skatt 'tax'	*hinder* 'obstacle'	*gur*	*ronung*
block 'block'	*värde* 'value'	*nold*	*sotter*
blyg 'shy'	*silver* 'silver'	*mind*	*lussin*
dans 'dance'	*vapen* 'weapon'	*hil*	*mektor*
dikt 'poem'	*manus* 'manuscript'	*bisk*	*pinter*
dygn 'day'	*kultur* 'culture'	*dul*	*galör*
grav 'grave'	*figur* 'figure'	*flatt*	*lensur*
krets 'circle'	*kalas* 'party'	*dalm*	*sental*
kund 'customer'	*metal* 'metal'	*nast*	*palang*
kvart 'quarter'	*panik* 'panic'	*töd*	*maket*
lind 'lime tree'	*rejäl* 'proper'	*haks*	*navott*

One third (ten monosyllabic and ten disyllabic words) had two IC. These are listed in the first ten rows of items in Table 1. Of the disyllabic words, nine had stress on the second syllable, and the remaining 21 had stress on the first syllable.

2.2 Stimuli – Nonsense words

The nonsense words were created by changing the first consonant of existing Swedish words. For instance, the nonsense word *ronung* was derived from the Swedish words *honung* 'honey' or *konung* 'king'. Each nonsense word could have been derived from at least two real words, so as to make the association between a nonsense word to any particular existing word as small as possible. For one third of the nonsense words, the first consonant was changed into a consonant that also occurred elsewhere in the word (e.g. *brand* 'fire' was changed into *drand*), for the other two thirds it was changed into a consonant that did not occur elsewhere in the word (e.g. *sorg* 'sorrow' was changed into *forg*). No nonsense word contained phonotactically illegal combinations of consonants or unlikely combinations of consonants and vowels.

For the rest, the composition of the list of nonsense words was the same as that of the real words. There were 30 monosyllabic and 30 disyllabic nonsense words. Nine disyllabic words had stress on the second syllable.

2.3 Practice items

Apart from the test stimuli, 20 practice items were created: ten monosyllabic, and ten disyllabic. Half of them were real words, the other half were nonsense words. The nonsense words were created in the same way as the test items. None of the practice items contained IC.

2.4 Recording and preparation of the stimuli

The stimuli were read by a female speaker of Swedish in a sound-proof studio. They were then digitized with a sample frequency of 16 kHz, and prepared for the experiment with the speech editor Praat.

2.5 Experimental procedure

The actual experiment was run in a sound-proof studio. The program was implemented on a Macintosh Power PC using Psyscope (Cohen et al. 1993). The order of the stimulus presentation was randomized for each subject. The stimuli were presented over headphones at a comfortable listening level.

In the beginning of the experiment, the subject received written instruction, after which he or she could ask questions in case anything was unclear. When all was clear, the subject sat down facing a computer screen and a button box placed on a table. The buttons on the button box were labelled 'real word' and 'nonsense word'. For half of the subjects the 'real-word' button was under their preferred hand, for the other half, the 'nonsense-word' button was under their preferred hand.

Table 2. Average error rates (%)

		real words	nonsense words
monosyllabic	with IC	3.5	2.0
	without IC	2.3	2.3
disyllabic	with IC	8.0	2.5
	without IC	1.3	3.0

The experiment started with the 20 practice items. Each stimulus was preceded by an exclamation mark on the computer screen in order to focus the subject's attention. After the practice items, there was a short break in the experiment in case the subject had any additional questions. After that, the actual test began, and all 120 test items were presented without any further breaks. The whole experiment took approximately 15 minutes per subject.

2.6 Subjects

In order to obtain a total group of 20 subjects, 23 subjects were tested. Three subjects were excluded for various reasons. All subjects were Swedish native speakers, eight men and twelve women. Most of them were students or staff of the language and literature departments at Lund University. None of them reported hearing problems. Their participation in the experiment was voluntary.

2.7 Dependent variables and statistical analysis

The reaction times (measured from stimulus offset) and the error rates were dependent variables. The data were analyzed with an F1–F2 analysis of variance.

3 Results

Table 2 shows the error rates for the experimental items. The overall error rates were lower than 3.5%, except for the disyllabic real words with IC. Inspection of the results revealed that the item *skakel* 'shaft' was responsible for the high error rate in this condition. Although this is an existing Swedish word, it mainly occurs in plural form in an idiomatic expression (*hoppa over skaklarna* which means 'to kick over the traces' or 'to be unfaithful'). In

Table 3. Average reaction times (ms)

		real words	nonsense words
monosyllabic	with IC	265	336
	without IC	246	427
	total	**252**	**396**
disyllabic	with IC	211	360
	without IC	195	392
	total	**200**	**381**

singular it has become rather old-fashioned, and therefore it was not recognized by many of the subjects. Since *skakel* was a debatable item it was excluded from further analysis. The overall error rates were so low that they were not analyzed statistically.

The average reaction times are listed in Table 3. Two effects were significant in the *F1* and the *F2*-analysis. First, it is a well-established finding that nonsense words take longer response times than real words. This was also the case in the present study. The responses to the real words were significantly faster than those to the nonsense words ($F1(1,19) = 51.780$, $p < 0.05$; $F2(1,111) = 75.008$, $p < 0.05$).

The second significant effect was the most important finding of the present study. There was a significant interaction of the lexical status of the items (real words or nonsense words) and the presence of IC ($F1(1,19) = 13.124$, $p < 0.05$; $F2(1,111) = 5.260$, $p < 0.05$). This interaction is shown in Figure 1. Real words with IC took longer time to accept than real words without IC, but nonsense words with IC took shorter time to reject than nonsense words without IC.

Looking at the figures in Table 2, it becomes clear that this overall pattern was clearer for the monosyllabic items than for disyllabic items. The difference between monosyllabic nonsense words with and without IC was 91 ms, whereas for the disyllabic nonsense words this difference was only 32 ms. Similarly, the difference between the monosyllabic real words with and without IC was 19 ms but for the disyllabic words it was 11 ms.

This overall difference between the monosyllabic and the disyllabic items was further reflected in three other effects that were only significant in the *F1*-analysis. First, the average reaction time to the monosyllabic words was longer (324 ms) than that to the disyllabic words (292 ms; $F1(1,19) = 8.515$, $p < 0.05$). Second, the average time to respond to items without IC was longer (315 ms) than that to the items with IC (295 ms; $F1(1,19) = 13.071$, $p < 0.05$). Third, there was an interaction of word length (monosyllabic or disyllabic) and the presence of IC ($F1(1,19) = 7.326$, $p < 0.05$). This interaction is represented in Figure 2. On the whole, monosyllabic items with IC took longer time to respond to than monosyllabic words without IC, but for the disyllabic words this difference was negligible.

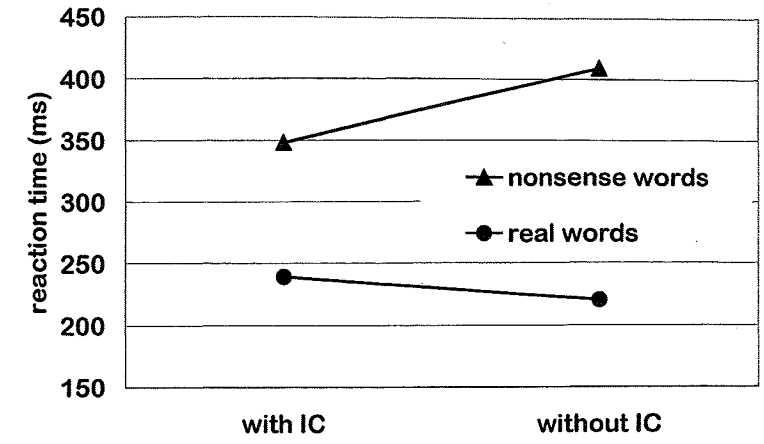

Figure 1. Interaction of lexical status and presence of IC

4 Discussion

In a previous study (van de Weijer 2003), it was shown that it is a relatively uncommon pattern to find two IC within the same monomorphemic word. The main purpose of the present study was to establish whether listeners have implicit knowledge of this aspect of phonological word structure. A lexical decision experiment was carried out in which subjects listened to nonsense words and real words in which the presence of IC was systematically varied.

The results revealed an asymmetrical pattern for the real words and the nonsense words, as was evidenced by a significant interaction of lexical status and the presence of IC. Real words without IC were recognized faster than real words with IC, whereas nonsense words without IC were rejected more slowly than nonsense words with IC. The pattern was the same for the

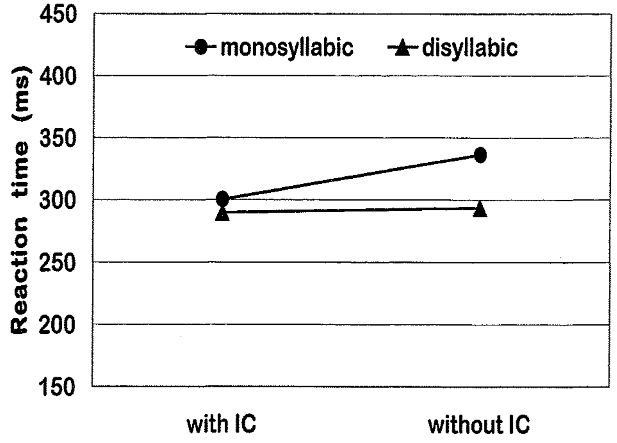

Figure 2. Interaction of word length and presence of IC

monosyllabic and the disyllabic items, but the differences were somewhat clearer for the monosyllabic items than for the disyllabic items.

This finding is clear support for the hypothesis that it is 'uncommon' or 'marked' to find two IC within a monomorphemic word. The listeners found it easier to reject nonsense words with IC, and found it more difficult to accept real words with IC. This result adds to the findings of the corpus study, that the consonants that a word is constructed of usually are different from each other. However, there are two alternative explanations that need to be ruled out.

The first alternative explanation is that the significant interaction is due to those items in which the IC were separated by a vowel only. In other words, it was the OCP effect that helped the listeners to reject the nonsense words with IC faster. This explanation cannot be ruled out completely by the present study. Furthermore, the pattern was clearer for the monosyllabic words than for the disyllabic words, suggesting that the distance between the IC does matter. However, in the majority of the test items the IC were separated by more than a single vowel (in most of the disyllabic items they were even separated by a syllable boundary). Moreover, a post hoc inspection

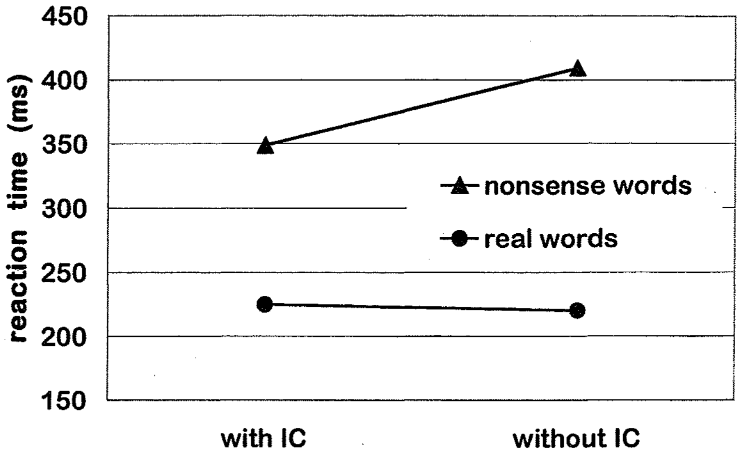

Figure 3. Interaction of lexical status and presence of IC for the items in which the IC were separated by more than a single vowel

of the results yielded the same pattern when all the items in which the IC were separated by a single vowel were excluded. This pattern is shown in Figure 3.

The difference between real words with and without IC is still there, even though it is smaller. And there is still a substantial difference between the nonsense words with and without IC.

A second possible alternative explanation is that the 'uniqueness point' at which the listeners could decide that an item was a nonsense item came earlier in the nonsense words with IC than in those without IC. In order to rule out this explanation, the nonsense words were checked with the dictionary, and the number of phonemes after which the listener could in principlce decide that it was a nonsense word was counted (e.g. *glagg* could have been *glans*, *glad*, etc., so that it could still be a real word up to three phonemes). The monosyllabic nonsense words with IC became unique after 2.8 phonemes on average and those with IC became unique after 2.9 phonemes on average. This is a very small difference which is unlikely to explain the difference in reaction times. For the disyllabic words, the difference was a bit larger. The uniqueness point of the disyllabic nonsense words with IC was 2.6 phonemes, and that of the disyllabic nonsense words without IC was 3.1 phonemes. This means that the difference in reaction times between the disyllabic nonsense words with and without IC possibly was due to the relatively early uniqueness point in the disyllabic nonsense words with IC. Nevertheless, this does not explain why the disyllabic real words with IC were accepted more slowly than those without. Moreover, the effect was stronger for the monosyllabic words than for the disyllabic words, so that it seems reasonable to assume that the differences were largely due to the presence of IC, and not to anything else.

How do the results contribute to our understanding of the speech comprehension process? The fact that there was an interaction of lexical status and the presence of IC clearly suggests that the listeners performed some kind of non-lexical analysis to the words, supporting the idea that listeners' sensitivity to phonological word structure plays a role in word recognition. The results do not imply that listeners use the investigated aspect of phonological word structure for lexical segmentation. The experimental task was not a segmentation task (as in the word-spotting experiments), but a recognition task. Nonetheless, the fact that monomorphemic words tend to be constructed of different consonants is a potentially useful source of information for segmentation. Two identical consonants may function as a signal for the listener that there is a word or a morpheme boundary. Whether this works as a cue for the listener is a question that is open for future experimental work.

A second question that needs to be addressed is *why* IC tend to be avoided within words. A remarkable fact, after all, is that consonant harmony is a common process in child language (Smith 1973). A number of words that were found in the corpus study (van de Weijer 2003) were words that are typically used by children (e.g. *mommy*, *daddy*, *nanny*, *cookie*, etc.). And yet, IC tend to be avoided in 'adult words'. The present study does not provide evidence for the reasons for varying the consonants, but the following two tentative explanations are proposed.

First, there might be an articulatory factor that plays a role. According to this explanation, it is easier to produce sequences with different consonants than sequences with the same consonants. Typically, tongue twisters tend to contain words with consonants or consonant combinations that are identical or similar. Interestingly, this does not appear to be the same for vowels. Languages with vowel harmony are relatively common, but languages with consonant harmony are rare.

A second explanation is that variation in the consonants within a word contributes to the coherence of the word. In order to clarify this explanation, consider grouping a series of randomly ordered coloured beads on a string. One strategy of doing this is to see when a specific colour comes back, so that each group contains that colour. Something similar could work for word recognition. Whenever a consonant is repeated, the listener assumes that a new group has begun. Admittedly, this is a speculative explanation, that needs extensive further investigation.

Finally, one way of finding out more about the reasons behind consonant variation within words and how the information may be used in speech perception is to examine whether the same pattern is found in other languages. Ongoing research suggests that this is indeed the case in other Germanic languages (Dutch, English and German). It will be even more interesting to analyze typologically different languages such as agglutinating languages, or languages with different phonological characteristics. These are projects for future research.

References

Clements, George & Elizabeth Hume. 1995. 'The internal organization of speech sounds'. In John Goldsmith (ed.), *The handbook of phonological theory,* 245-306. Oxford: Basil Blackwell.

Cohen, Jonathan, Brian MacWhinney, Matthew Flatt & Jefferson Provost. 1993. 'PsyScope: A new graphic interactive environment for designing psychology experiments'. *Behavioral Research Methods, Instruments, and Computers* 25:2, 257-71.

Cole, Ronald & Jola Jakimik. 1980. 'A model of speech perception'. In Ronald Cole (ed.), *Perception and production of fluent speech,* 133-64. Hillsdale, NJ: Erlbaum.

Cutler, Anne & David Carter. 1987. 'The predominance of strong syllables in the English vocabulary'. *Computer Speech and Language* 2, 133-42.

Cutler, Anne & Dennis Norris. 1988. 'The role of strong syllables in segmentation for lexical access'. *Journal of Experimental Psychology: Human Perception and Performance* 14, 113-21.

Jusczyk, Peter, Anne Cutler & Nancy Redanz. 1993. 'Infants' preference for the predominant stress patterns of English words'. *Child Development* 64, 675-87.

Jusczyk, Peter, Angela Friederici, Jeanine Wessels, Vigdis Svenkerud & Ann Marie Jusczyk. 1993. 'Infants' sensitivity to the sound patterns of native language words'. *Journal of Memory and Language* 32, 402-20.

Jusczyk, Peter, Paul Luce & Jan Charles-Luce. 1994. 'Infants' sensitivity to phonotactic patterns in the native language'. *Journal of Memory and Language* 33, 630-45.

McQueen, James. 1998. 'Segmentation of continuous speech using phonotactics'. *Journal of Memory and Language* 39, 21-46.

McQueen, James, Dennis Norris & Anne Cutler. 1994. 'Competition in spoken word recognition: spotting words in other words'. *Journal of Experimental Psychology: Learning, Memory and Cognition* 20, 621-38.

Newsome, Mary & Peter Jusczyk. 1995. 'Do infants use stress as a cue in segmenting fluent speech?' In Dawn MacLaughlin & Susan McEwen (eds.), *Proceedings of the 19th Boston University Conference on language development*. Boston, MA: Cascadilla Press.

Smith, Neil. 1973. *The acquisition of phonology*. Cambridge University Press.

Weijer, Joost van de. 2003. 'Consonant variation within words'. In Dawn Archer, Paul Rayson, Andrew Wilson & Tony McEnery (eds.), *Proceedings of the Corpus Linguistics 2003 Conference* (University Centre for Computer Corpus Research on Language Technical Papers, vol. 16), 184-90. Lancaster University, United Kingdom.

Joost van de Weijer <vdweijer@ling.lu.se>

Lund University, Dept. of Linguistics
Working Papers 51 (2005), 239–255

A case study of impersonation from a security systems point of view

Elisabeth Zetterholm, Daniel Elenius and Mats Blomberg

1 Introduction

Impersonation of a person, especially by means of voice, is sometimes used to amuse a human audience. Imitations often sound quite convincing. For several reasons it would be interesting to establish what aspects are important in performing a successful impersonation act. There are several components besides the acoustic signal that contribute to the subjective impression, such as the mood, the expectation, etc., and, if the impersonator can be seen, also visual similarity and gestural patterns. It may be difficult to determine the contribution of each individual factor.

It would also be interesting to use a more objective measure of performance than human impression. Objective measures of many of these aspects may be quite difficult to extract due to their complexity and lack of standardised analysis techniques. The acoustic signal, however, is suitable for objective evaluation, since there exist established techniques for quantitative phonetic analysis of speech and algorithms for determining the degree of acoustic similarity between utterances spoken by different persons.

Spectral analysis has been used by Zetterholm 2003, who showed that, for instance, the professional impersonator adjusted his fundamental frequency and the formant frequencies of the vowels during impersonation to be closer to the target voice compared to that of his natural voice. In the present report, we complement these measurements with a computer-based speaker verification system. This type of system is normally used to judge, by the acoustic properties of a spoken utterance, whether a person has the identity (s)he has claimed or not. Our idea behind using this system is to measure how close to the target voice a professional impersonation might be able to reach and to relate this to phonetic-acoustic analysis of the mimic speech.