Conflicting acoustic cues in stop perception

Rolf Carlson
KTH, CSC, Dept. Speech, Music and Hearing

Introduction

In this paper we will return to a basic question in phonetics: the relation between acoustic cues and perceived phonetic units. The research has been focused on the perception of unvoiced stops. The rationale for conducting the experiments is to further illuminate and support the hypothesis that speech perception is a dynamic and adaptive perceptual process in the interpretation of acoustic cues.

A special inspiration for the research has been the increased interest in fine phonetic details and the studies of temporal integration in speech perception (Hawkins, 2003). An unpublished pilot experiment, carried out during the 70s by Gunnar Fant, gave an intriguing illustration of the active processing in speech perception. The third formant in a natural front vowel was moved with the help of a pole-zero filter resulting in a perceptual vowel shift. However, if a sequence of different vowels were filtered with this stationary setup the perceived vowel identity was not shifted. The perceptual process was able to identify the filtering as a distortion and disregard that a formant was misplaced.

Already the classical experiments on the perception of stops (Liberman et al., 1952) are looking for universal cues that can be used as robust features in speech perception. The results are proposed to point to an invariance that should be search for on an articulatory level rather than an acoustic one. However, the persistent work by Stevens on invariant cues points to how acoustic landmarks and features are used in human perception (e.g. Stevens, 2002). In the perceptual experiments by Carlson et al. (1972) the search for acoustical cues is carried out by replacing segments, sometimes even filtered ones, in speech waveforms. A review of speech perception including several studies on stops has recently been published by Hawkins (2004).

Experiment

A straightforward waveform splicing technique has been used to create stimuli with sometimes contradicting acoustical cues. Using three classical manipulations we form a baseline for the fourth type of stimuli, where we evaluate if a temporal context can reduce the perceptual impact of stop release acoustic cues.

Clearly spoken nonsense words by one speaker were selected as original stimuli. Eighteen nonsense words /te_'C_V_de/ were used including one of three unvoiced stop consonants (p,t,k) before one of six vowels (a, a:, i, i:, u, u:).

In the stimulus type initial, the first syllable in each original word /te_C_V_de/ was replaced by the corresponding part of another stimulus word. The inserted segment came from a word with a different consonant C but the same vowel V. The mixing point was placed in the stop gap of the consonant C.

In the release type of stimulus only the stop release in C (40 ms) in a word was replaced by another equally long stop release. The new segment came from a word with a different consonant C but the same vowel V. A few samples before and after the mixing points were interpolated to avoid an unwanted distortion.

The combination of the two last types form the initial+release type replacing the first part of a word with another corresponding part. As a result only part of the aspiration of the original consonant C and the two final syllables are kept intact from the original word

Finally, the repeated release type stimuli has the same processing as the release type but in addition the replacing release in consonant C is repeated at regular intervals. The sequence of repeated releases creates a distortion in the stimuli. It is important to remember that one replacing release is still at the same position in consonant C as in the release type, see Figure 1.

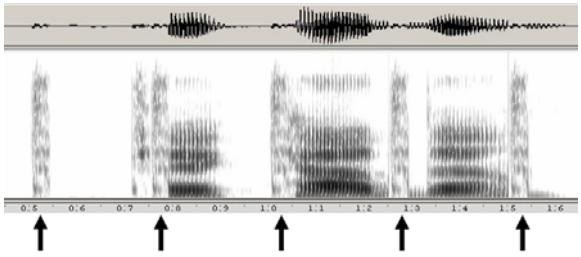

*Figure 1. Example of a repeated release stimulus. The release (40ms) in "te'**k**adde" has replaced the release in "te'**t**adde" (middle arrow). Furthermore, the k-release is repeated at regular intervals (marked with arrows).*

Using a web interface, seven subjects were instructed to simply report which unvoiced stop (p, t, k) they heard in the middle of the respective nonsense word. Perceptual data for the individually randomized sequence of the stimuli was collected.

Results

We will in the following use the abbreviation **COP** (change of percept) when we discuss the perceptual results. An example of such a change is when a p-release replacing a t-release in /te'tade/ changes the perception of the word to /te'pade/. The COP results grouped according to stimulus type and vowel length are presented in Figure 3.

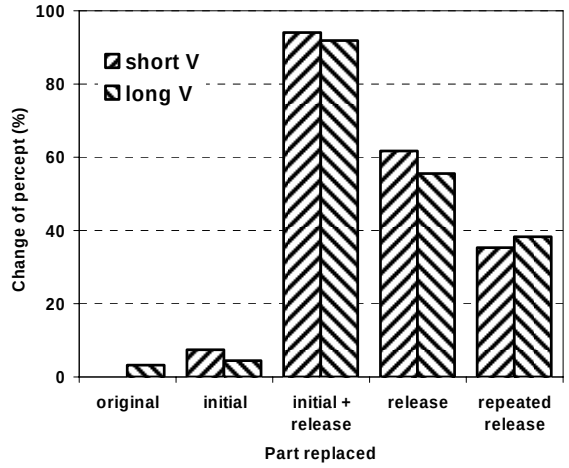

Figure 2. Change of percept (COP) grouped according to stimulus type and vowel length.

As expected the *initial+release* type has a very high COP (93%), while the *initial* type has very little impact on the consonant identity (6%). More than half of the stimuli changed their identity for the *release* type (59%), where only the burst segment was replaced. Finally, the *repeated release* type has a COP of only 37% compared to the 59 % COP for the *release* type.

Discussion

The COP results for the *initial* type and the *release* type are according to what can be predicted based on general knowledge of speech perception and the publications reviewed in the introduction. The cues in the proceeding vowel are weaker than the cues in the stop release. However, it is interesting to note that the combination of acoustical cues in the proceeding vowel and the stop release generates a stronger COP than a simple addition of the results from the individual sets of cues. One could then speculate that the cues in the proceeding vowel play an important secondary role by adding robustness to the percept. A coherence of the acoustic cues makes the decision less confusing for the listener.

How the subjects perceived the *repeated release* type stimuli compared to the *release* type is the other focus of the experiment. By repeating a sequence of releases we introduce an unnatural distortion. The COP results suggest that listeners correctly classified the repeated releases as a distortion and thus tried to disregard the disturbing acoustic releases in the identification process. Unfortunately for the listener this also applied to the correctly aligned stop release. Thus, the replacing stop release cues received less importance compared to the same replacement in the release type of stimuli.

Conclusion

By the use of repeated releases we show that a temporal distortion can have the same kind of impact on a percept as a spectral distortion. If the manipulation is classified as a disturbance, the cues have reduced importance for the classification. Furthermore, the result shows that acoustic cues in a vowel preceding a stop can have a significant influence on the percept if they are in accordance with the stop release cues.

References

Carlson R, Granström B & Pauli S (1972). Perceptive evaluation of segmental cues. In Proceedings of the Conference on Speech Communication and Processing (pp. 206-209). Bedford, MA, USA. also STL-QPSR, 13(1), 018-024.

Hawkins S (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics* 31: 373-405.

Hawkins S (2004). Puzzles and patterns in 50 years of research on speech perception. *From Sound to Sense: 50+ Years of Discoveries in Speech Communication*, MIT, Cambridge, MA, USA

Liberman A, Delattre P & Cooper F S (1952). The role of selected stimulus-variables in the perception of unvoiced stop consonants, *Am. J. of Psych*., 497-516.

Stevens K N (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, 111, 1872–1891.