Phonological and Phonetic Representations: The Case of Neutralization

Allard Jongman
University of Kansas

1. Introduction

The present paper focuses on the phenomenon of phonological neutralization to consider the role of phonological and phonetic representations in the production and perception of speech. Phonological neutralization involves the elimination of a phonemic distinction in a particular phonological context. Neutralization of a word-final voicing contrast is a classic example. Table 1 shows the relevant forms for German:

Orthographic form	Gloss	Phonological (underlying) form	Phonetic (surface) form
Rat	advice	/ʁat/	[ʁat]
Rad	wheel	/ʁad/	[ʁat]

Table 1. Minimal wordpair with distinct phonological representations but, putatively, identical phonetic representations.

While Rat and Rad differ in their underlying representation in terms of the voicing of the final consonant, the surface forms are identical, both ending in a voiceless alveolar stop, according to phonological analysis. In the past 20 years, phonetic research has attempted to document the extent to which this neutralization is phonetically complete. If neutralization is indeed phonetically complete, the two surface forms should be phonetically identical. If neutralization is incomplete, the two surface forms do not only differ but they should differ in predictable ways. That is, the acoustic correlates should be of the same quality as when the distinction is fully maintained.

Most phonetic research on neutralization has concentrated on word-final devoicing, particularly in languages such as German, Polish, and Catalan (e.g., Charles-Luce, 1985; 1993; Dinnsen and Charles-Luce, 1984; Fourakis and Iverson, 1984; Jassem and Richter, 1989; Port and Crawford, 1989; Port and O'Dell, 1985; Slowiaczek and Dinnsen, 1985). Some of these studies report incomplete neutralization while others document complete neutralization. Factors known to influence neutralization and therefore potentially responsible for these differences across studies include orthography (whether or not the underlying distinction is represented in the spelling, Fourakis and Iverson, 1984), speech style (careful vs. casual, Port and Crawford, 1989), and semantic expectancy (whether or not the context is predictive of the target form, Charles-Luce, 1993).

2. Experiment 1: Manner neutralization in Korean

Since most research has concerned neutralization of a voicing distinction and results vary across studies, it is not clear a priori what one would expect to find for neutralization of a qualitatively different type. In order to address that issue, Hyunsoon Kim and I explored neutralization of manner of articulation in Korean (see Kim and Jongman, 1996, for a detailed report). The relevant forms are shown in (1):

(1) /kət/	"to collect"	/kət + ə/	[kə.də]	(with indicative marker /ə/)
		/kət + tʃi/	[kət.tʃ'i/	(with negative marker /tʃi/)
/kət^h/	"outside"	/kət^h + ɨl/	[kə.t^{h}ɨl]	(with object marker /ɨl/)
		/kət^h + kwa/	[kət.kwa]	(with marker /kwa/ "and")
/kəs/	"outside"	/kəs + i/	[kə.si]	(with subject marker /i/)
		/kəs + kwa/	[kət.kwa]	(with marker /kwa/ "and")

These three forms differ in their underlying representation. However, when followed by a suffix beginning with a consonant, all three forms surface with a plain voiceless [t]. In order to determine whether this neutralization of manner of articulation is phonetically complete, we focused on two parameters, vowel and final consonant duration. In particular, in the case of incomplete neutralization we expected to find a longer vowel duration in forms that underlyingly end in /s/ as compared to /t, t^h/. This is because in Korean, like many other languages, vowels are longer when preceding fricatives relative to stops. In addition, we expected final consonants derived from underlying /t^h/ to be longer than those derived from /t/.

Tables 2 and 3 illustrate the pattern of results that we obtained:

	Underlying stem-final consonant		
Speaker	**/t/**	**/t^h/**	**/s/**
1	81	79	82
2	75	76	73
3	80	84	85
4	58	57	53

Table 2. Mean vowel duration (in ms) for each speaker, for minimal triplet members underlyingly ending in /t/, /t^h/, and /s/.

	Underlying stem-final consonant		
Speaker	**/t/**	**/t^h/**	**/s/**
1	81	79	82
2	75	76	73
3	80	84	85
4	58	57	53

Table 3. Mean consonant duration (in ms) for each speaker, for minimal triplet members underlyingly ending in /t/, /t^h/, and /s/.

There were no differences in vowel or consonant duration as a function of the manner of articulation of the underlying consonant. However, before claiming that these results demonstrate that neutralization of manner in Korean is phonetically complete, it is important to realize that we measured only two acoustic cues. It is therefore possible that differences due to underlying form would show up in some other parameter that we did not measure. Since the number of acoustic cues that would have to be measured to address this issue is potentially very large, we decided to conduct a perception experiment instead to determine whether we had overlooked any cues that allowed listeners to tell the underlying forms apart.

Korean listeners were presented with a surface form (e.g., [kət] derived from either /kət/, /kət^h/, or /kəs/) and asked to indicate which word they heard. Overall identification accuracy was at 32%, virtually identical to chance performance. Thus, these forms did not contain any cues that allowed listeners to decide which word they had heard. Taken together, these acoustic and perceptual results suggest that Korean manner neutralization is indeed phonetically complete. In view of some previous reports of incomplete neutralization of voicing, our finding of complete neutralization of manner raises the question whether these different results are due to the fact that neutralization of voicing and manner are qualitatively very different or to differences in the methodology employed. We therefore decided to conduct our own investigation of voicing neutralization.

3. Experiment 2: Voicing neutralization in Dutch

Similar to German, Dutch has a process of word-final voicing. The relevant forms are shown in Table 4 (plurals are included to illustrate that the underlying consonant does surface):

Orthographic form	Gloss	Phonological (underlying) form	Plural	Phonetic (surface) form
pond	pound	/pɔnd/	/pɔndən/	[pɔnt]
pont	ferry	/pɔnt/	/pɔntən/	[pɔnt]

Table 4. Minimal wordpair with distinct phonological representations but, putatively, identical phonetic representations.

We analyzed minimal pairs produced by 15 speakers (a large corpus for this kind of neutralization study) by measuring vowel duration, consonant closure duration, and burst duration (see Warner, Jongman, Sereno, and Kemps, in press, for details). Results (averaged across words with phonemically short and long vowels) are summarized in Table 5:

	Underlying word-final consonant	
	Voiceless	Voiced
Vowel duration	148	152
Burst duration	135	129
Closure duration	78	76

Table 5. Duration (in ms) as a function of the underlying voicing of the word-final consonant.

Statistical analysis indicated that the differences for vowel and burst duration were significant. While these differences are numerically quite small, their significance is presumably due to the large number of speakers and words employed.

We then explored whether listeners are sensitive to these differences. Instead of including all words produced by all speakers, we selected four speakers differing in the way they had produced the minimal pairs. Table 6 shows if and how each speaker used vowel and burst duration to distinguish the underlyingly voiced and voiceless tokens.

Speaker	Vowel duration	Burst duration
A	++	++
B	++	+
C	0	-
D	-	-

Table 6. Extent to which each speaker used vowel duration and/or burst duration to distinguish underlyingly voiced and voiceless tokens. ++ and + indicate a relatively large or small difference, respectively, in the expected duration , 0 indicates no difference, and - indicates a small difference in the opposite direction (see text).

For example, Speaker A had relatively large differences in the expected direction for both vowel and burst duration. In other words, this speaker produced a substantially longer vowel and shorter burst for forms ending in an underlyingly voiced consonant. Speaker C did not differentiate forms in terms of vowel duration but produced a *longer* burst duration for forms ending in an underlyingly voiced consonant. In order to maximize our listeners' chances, the stimuli were blocked by speaker. The results are shown in Table 7:

Speaker	Intended final /t/	Intended final /d/
A	63*	50
B	62*	50
C	54	55
D	57	61*

Table 7. Percent 't' responses as a function of whether the speaker produced a form ending in underlying /t/ or /d/. * indicates a significant difference between the two response rates.

As can be seen from Tables 6 and 7, these results show a very good match between production and perception. Words with final /d/ and /t/ produced by speakers A and B who distinguished the voiceless and voiced forms in terms of vowel and closure duration were correctly identified at significantly better than chance levels. For example, for Speaker A, there were significantly more 't' responses to words underlyingly ending in /t/ (63%) than to words underlyingly ending in /d/ (50%). Identification rates for Speaker C, who did not distinguish the different underlying forms in terms of vowel duration and only slightly so - but in the opposite direction - for closure duration were not significantly different. Finally, Speaker D distinguished the different underlying forms by using vowel and closure duration in the opposite direction and, lo and behold, perception exactly mimicked this pattern in that listeners perceived significantly more 't' responses when the speaker produced words ending in /d/!

We also conducted two additional perception experiments that went beyond the kind of perception experiment that is typically done in neutralization studies (as in Kim and Jongman, 1996). Both of these required the creation of duration continua by digitally editing the natural productions. In one experiment, we manipulated the cue that seemed most salient based on our findings so far. A look at Table 7 suggests that vowel duration is the most likely cue to the underlying voicing distinction. We therefore created vowel duration continua for several words produced by each of the four speakers used previously. For example, an 11-step continuum was created from /wɑt/ to /wɑd/ with the original production (either /wɑt/ or /wɑd/) in the middle and a vowel duration range of 50 ms. Stimuli were blocked by speaker and continuum and listeners were asked to respond with the word they thought the stimulus sounded more like. This is a difficult task given that vowel duration varied over only a small range (based on our production data) and the final consonant did not vary at all. Nevertheless, the results revealed that listeners were able to used vowel duration as cue. They gave significantly more 't' responses for short vowel durations as compared to long vowel durations.

Interestingly, we obtained very similar results in the other perception experiment in which consonant closure duration continua were created. As shown in Table 5, consonant closure duration did not systematically vary as a function of underlying consonant voicing. One may therefore expect that it does not constitute a reliable perceptual cue. However, for our continua with the original word in the middle and a closure duration range of 100 ms, listeners gave significantly more 't' responses for longer closure durations.

Taken together, the results from these two perception experiments in which acoustic parameters were systematically manipulated indicate not only that listeners can use vowel duration as a cue but they can also use closure duration as a cue even though it does not serve as a cue in natural speech. Apparently, listeners identify the only possible perceptual cue (i.e., the only property that varies within the confines of the experiment) and apply it based on its relevance in a different environment.

A number of years ago, Dan Dinnsen, one of the primary researchers in the area of phonetic and phonological neutralization, published an inventory of possible types of neutralization, taking into account both the production and perception domains. The major features of this taxonomy (Dinnsen, 1985) are reproduced here as Table 8:

Type	Production differences	Perception differences	Comments and examples
A	No	No	Standard view Not well established
B	Yes	No	Sound change in progress
C	Yes	Yes	Non-neutralizing German devoicing
D	No	Yes	Impossible

Table 8. Four logically possible types of neutralization when considering both production and perception (after Dinnsen, 1985).

The results of our research on Korean and Dutch indicate the need for some modifications to this table. First, Dinnsen's claim that Type A neutralization, the standard phonological view, is not well-established is challenged by our finding that neutralization of manner of articulation in Korean is complete in terms of both production and perception. Similar findings have also been obtained for the perception and production of Japanese pitch accent (Maniwa and Jongman, submitted). Types B and C are well established and presumably distinguished by the magnitude of the acoustic difference. If the difference is reliable but too small to be perceived, Type B neutralization is the result. Type C occurs when the difference is large enough to be perceived. Type D, is, of course, logically impossible. However, the results from the perception experiment on Dutch closure duration add a twist in that listeners can be made to use a cue that does not vary at all in production.

Finally, any discussion of phonological and phonetic representations needs to address the psychological reality of such representations. The next experiment attempted to specifically determine if listeners make use of phonological representations when perceiving speech.

4. Experiment 3: Phonological representation of [voice] in speech perception

This experiment involved the processing of Dutch verb+clitic constructions (for details, see Lahiri, Jongman, and Sereno, 1990; Jongman, Sereno, Raaijmakers, and Lahiri, 1992). In colloquial Dutch, when the clitic 'der' ("her") attaches to a preceding verb ending in an obstruent, the verb+clitic combination leads to an optional voicing alternation on the surface. Relevant forms are shown in Table 9:

Verb	Underlying stem	Gloss	First ps. Sg.	Gloss	Verb+clitic (voiceless)	Verb+clitic (voiced)
[kœsən]	/kœs/	'to kiss'	[ɪk kœs haɹ]	'I kiss her'	[ɪk kœstəɹ]	[ɪk kœzdəɹ]
[kizən]	/kiz/	'to choose'	[ɪk kis haɹ]	'I choose her'	[ɪk kistəɹ]	[ɪk kizdəɹ]

Table 9. Verb stems and voiceless and voiced verb+clitic constructions.

The verb stems for 'to kiss' and 'to choose' end in a voiceless and voiced obstruent, respectively. Crucially, for each underlying verb stem, two clitic forms can surface, one that matches and one that mismatches the underlying stem in terms of voicing.

In order to determine whether the parsing of the cliticized forms and recognition of the verbs was affected by either surface phonetic or underlying phonological representations of the verb stem, a primed auditory lexical decision paradigm was used. As illustrated in (2), listeners would hear a target (a verb stem or infinitive form, they are identical in Dutch) preceded by a prime that was a verb+clitic construction that either matched or mismatched the target in terms of its underlying voicing representation.

(2) *underlying stem-final consonant* *prime* *target*

/s/	[ɪk kœstəɹ] [ɪk kœzdəɹ]	[kœs]
/z/	[ɪk kistəɹ] [ɪk kizdəɹ]	[kis]

Verb stems ending in /p, b/ were used in addition to stems ending in /s, z/, The same listener did not hear both the voiceless and voiced prime of the same verb. Results are shown in Figure 1.

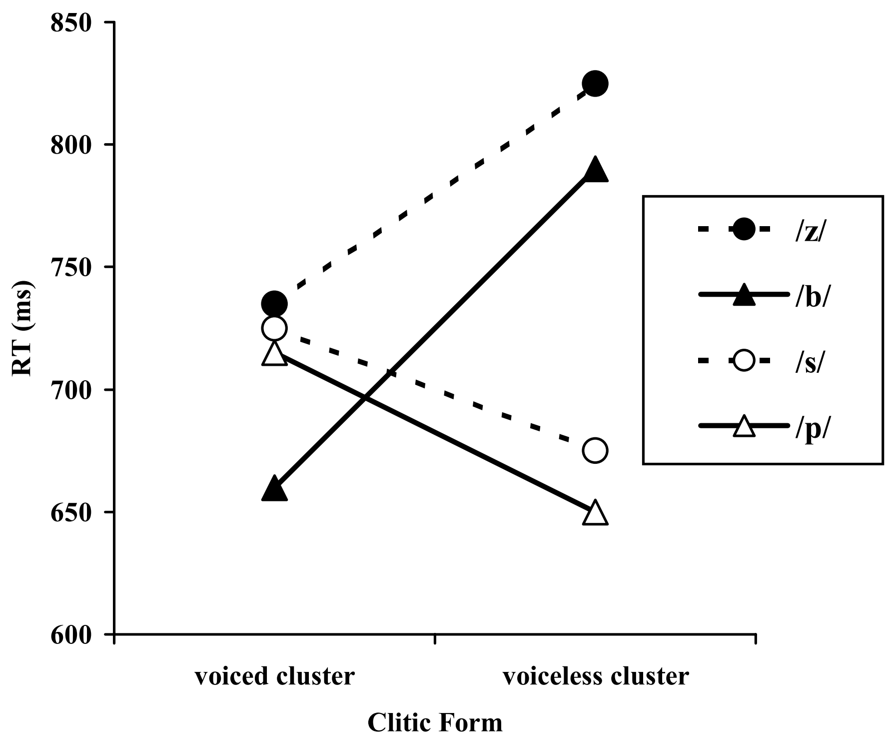

Figure 1. Reaction times (in ms) to voiced and voiceless clitic forms as a function of their underlying stem-final consonants (/p, b, s, z/).

These data clearly show an asymmetry in response latencies to the same target verb depending on whether or not the listener heard the surface form that matches the phonological representation of that verb. For verbs underlyingly ending in a voiceless obstruent, responses were faster when preceded by the voiceless clitic form. Conversely, when verbs ended in voiced obstruent, responses were faster when they were preceded by the voiced clitic form. This suggests that it is not the case that both variants of a verb form are stored in the lexicon. Instead, recognition is influenced by the underlying phonological representation of the verb stem.

5. Conclusions

The Korean data suggest that neutralization of manner of articulation is phonetically complete and as such support the standard view in phonological theory. In contrast, the Dutch data indicate that neutralization of voicing is phonetically incomplete. This is in agreement with some previous research on German, although the Dutch durational differences are much smaller than what has been reported for German. One possible motivation for incomplete neutralization might be perceptual in nature. After all, a distinction that is not completely neutralized may be easier to recover. However, that of course depends again on the magnitude or perceptual salience of the distinction and as Dinnsen (1985) already pointed out, cases have been documented in which speakers make systematic distinctions but below the level at which listeners can profit from them. In addition, the Dutch data show the reverse pattern, namely that listeners can use differences that are not reliably produced.

Assuming that both the Korean and Dutch studies have been carefully conducted, we now have solid evidence for complete neutralization in one and incomplete neutralization in the other. Both studies were quite similar in design, and both languages represent the underlying representation in their spelling, to mention two obvious reasons for the differences observed. The question then becomes why only *some* distinctions are completely neutralized. Possible answers include the following: some distinctions may be more categorical than others. For example, the distinction between a fricative and a stop can be considered more categorical than that between a voiced and voiceless obstruent. This would account for the finding of complete neutralization of manner but incomplete neutralization of voicing. Differences could also be due to the fact that underlying distinctions are cued by multiple acoustic parameters. In this view, degree of neutralization is determined by the extent to which each parameter can be successfully neutralized. Differences may also be caused by the distinction between lexical and postlexical processes, or by paradigmatic pressure (see Mascaro, 1987).

Finally, differences may arise from the fact that phonetic implementation rules might refer to more than one representation. That is, the lexicon could have multiple underlying forms. In her insightful discussion of neutralization, Blumstein (1991) points out that if neutralization were always complete, the phonetic evidence would be consistent with the phonological analysis but would be unable to *confirm* it. Specifically, phonetic evidence could not rule out a lexicon with multiple underlying forms. For example, referring to Table 5, even though the singular-plural alternation is regular and predictable, the word 'pond' could have two underlying forms, /pɔnt/ for the singular and /pɔnd/ for the plural. However, incomplete phonetic neutralization rules out such a view. The results from the experiment with Dutch cliticized forms also suggest that a single, rather than multiple, forms are stored. In sum, then, rather than calling the concept of linguistic contrast into question, incomplete neutralization in fact supports the standard phonological analysis of neutralization.

References

Blumstein, S.E. (1991). The relation between phonetics and phonology. *Phonetica 48*, 108-119.

Charles-Luce, J. (1985) Word-final devoicing in German: Effects of phonetic and sentential contexts, *Journal of Phonetics,* **13**, 309-324.

Charles-Luce, J. (1993) The effects of semantic context on voicing neutralization, *Phonetica,* **50**, 28-43.

Charles-Luce, J. & Dinnsen, D. (1987) A reanalysis of Catalan devoicing, *Journal of Phonetics,* **15**, 187-190.

Dinnsen, D. (1985) A re-examination of phonological neutralization, *Journal of Linguistics,* **21**, 265-279.

Dinnsen, D. & Charles-Luce, J. (1984) Phonological neutralization, phonetic implementation and individual differences, *Journal of Phonetics,* **12**, 49-60.

Fourakis, M. & Iverson, G. (1984) On the 'incomplete neutralization' of German final obstruents, *Phonetica,* **41**, 140-149.

Jassem, W. & Richter, L. (1989) Neutralization of voicing in Polish obstruents, *Journal of Phonetics,* **17**, 317-325.

Jongman, A., Sereno, J., Raaijmakers, M., and Lahiri, A. (1992). The phonological representation of [voice] in speech perception. *Language and Speech 35,* 137-152.

Kim, H., and Jongman, A. (1996). Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean. *Journal of Phonetics 24,* 295-312.

Lahiri, A., Jongman, A., and Sereno, J. (1990). The pronominal clitic [dər] in Dutch: A theoretical and experimental approach. In G.E. Booij and J. van Marle (eds.), *Yearbook of Morphology, Vol. 3* (pp. 115-127). Dordrecht: Foris Publications.

Mascaro, J. (1987). Underlying voicing recoverability of finally devoiced obstruents in Catalan. *Journal of Phonetics 15,* 183-186.

Port, R. & Crawford, P. (1989) Incomplete neutralization and pragmatics in German, *Journal of Phonetics,* **17**, 257-282.

Port, R. & O'Dell, M. (1985) Neutralization of syllable-final voicing in German, *Journal of Phonetics,* **13**, 455-471.

Slowiaczek, L. & Dinnsen, D. (1985) On the neutralizing status of Polish word-final devoicing, *Journal of Phonetics,* **13**, 325-341.

Warner, N., Jongman, A., Sereno, J., and Kemps, R. (in press). Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch. *Journal of Phonetics.*