SIMILARITY AND NEWNESS - WORKABLE CONCEPTS IN DESCRIBING PHONETIC CATEGORISATION?

Duncan Markham

Department of Linguistics and Phonetics
Lund University, Sweden

ABSTRACT - Experimental and impressionistic data from learners of the sounds of Japanese and English are presented and discussed with regard to perceptual processing and a model of speech sound learning (Flege's SLM). It is argued that the perceptual classifications 'similarity' and 'newness' as proposed by Flege are not workable, and some alternative characteristics and criteria used in sound analysis by learners are posited.

INTRODUCTION

The literature on pronunciation and phonetic acquisition is considerable, yet we still have little concrete understanding of why certain second language (L2) sounds are harder to acquire than others for speakers of a variety of different first languages (L1s). One model which has been proposed to directly address this problem is discussed in light of L2-acquisition data and informal observations of learners.

Flege's Speech Learning Model (Flege, 1992) uses comparative classifications for determining how a learner will react to and ultimately acquire sounds in an L2. Sounds are classed as 'new', 'similar', or 'identical' on the basis of the difference between L2 sounds and existing L1 sounds, and the model then predicts how the learner will behave. Sounds which are classed as being *similar* to an existing L1 sound will only be produced as the L1 sound and never as an authentic L2 sound, whereas sounds judged to be *new* will be learnt successfully (in a native-like manner) by learners. *Identical* sounds will present no problem for the learner, as all necessary articulatory knowledge is already available in the L1. Although these three classifications are intended to be *predictive*, the criteria used are underdefined. Flege's only proposal for criteria of which I am aware is based on comparison of IPA transcriptions (Flege, op cit). Sounds in two languages which are broadly transcribed using the same IPA symbol are posited as being similar, and those transcribed with different symbols are posited as being new. I suspect, however, that most phoneticians would readily be able to think of examples which falsify Flege's claims. Furthermore, the Speech Learning Model seems to apply without regard to deep *phonological* or *phonotactic* interference from the learner's L1, and makes no allowance for individual phonetic experience.

Although Flege has written much on this topic, and presented data which he claims to support the model, few other researchers have directly tested it. The two independent studies I have identified which directly address Flege's claims, or at least relevant issues, are Polka (1991) and Munro (1993). Polka tested a group of American English speakers on their perception of a series of stop contrasts in Hindi. The informants' perception of dental and retroflex prevoiced, breathy voiced, voiceless unaspirated, and voiceless aspirated stops could not be predicted on the basis of knowledge of the structure of English, contrary to Flege's claims. Post-test acoustic analyses of the stimuli revealed correlations between perceptual behaviour and certain *acoustic* characteristics of the stops. The features corresponded to both perceptually redundant and perceptually significant acoustic characteristics of English.

Munro (1993) looked at the pronunciation of English vowels by a group of Arabic speakers of English. He observed that the speakers produced certain vowels very well (based on independent accent judgement) and other vowels poorly. This corresponds to the expectations of the Speech Learning Model. However, the subjects' good vowels were those which the SLM would predict to be the hardest to learn, whilst their poor vowels were those which theoretically should have been easiest. The SLM's predictive power would thus seem to be somewhat questionable.

In this paper I address the viability of newness and similarity as processing concepts, and whether these predictive criteria can be quantified.

As Munro's speakers had had varying experience with English (between 1 and 27 years), and were tested on material which was fairly simple (bVt/bVd syllables), data from an experiment using more

complex stimuli and totally naive speakers were taken from a corpus of pronunciation material and are presented here.

EXPERIMENTAL DATA

The data to be presented come primarily from a group of eight high-ability language learners, with Swedish as their L1. These subjects imitated words and phrases in various contexts for a number of languages to which they had had no prior exposure. (for detailed methodology, see Markham, 1995). Most of the examples for this paper come from the subjects' imitations of Japanese stimuli. Additional examples are taken from informal observations of Swedish and German speakers of English.

The subjects were fitted with lightweight headset microphones and high-quality headphones. After having been familiarised with the stimulus speaker's voice (Osaka Japanese) and having heard two (slow, normal) readings of a text in Japanese, the subjects heard the stimuli for imitation. The learners were not provided with any phonetic representation of Japanese (ie, neither romaji, hiragana/katakana, nor phonetic transcription), thus having to rely solely on auditory input. Each stimulus phrase was heard three times at each speed, and then the individual stimuli for imitation were presented. The stimuli were at first single words from the phrase and gradually became more complex (subphrases and then whole phrases). Each stimulus presentation consisted of two instances of the item, which was then imitated by the subject. A total of three presentations were made for each stimulus, yielding three imitations per subject per stimulus.

The stimuli were the following three phrases, and words and subphrases contained therein:

sutefan wa pairotto de
/sɯte[ɸ]an wa pairoto:o de/
[sᵚtɐɸãⁿ ɤ̆a pairoto:o de]

ofu to itta guaidesu
/ohɯ to it:a gɯaidesɯ/
[oɸɯ̆ to it:a gɤ̆ãidesᵚ]

futariwa Nihon Kookuu de hataraite imasu
/hɯtariwa nihon ko:kɯ: de hataraite imasɯ/
[ɸᵚtariɤ̆a nihõⁿ ko:kɯ̆: de hataraite imasᵚ]

Four sounds in the Japanese material were selected for examination here:

1a. a 'similar' sound [ɸ] (voiceless bilabial fricative),
1b. a 'similar' sound [ɾ] (voiced alveolar tap),
2. an 'identical' sound [ç] (voiceless alveolopalatal fricative) (this sound taken from a set of stimuli not shown above),
3. and a 'new' sound [ɤ̆] (bilabial approximant).

RESULTS

predicted perception	*Swedish inventory*	***Japanese stimulus***	*imitators' output*
1a similar	[ɧʷ, f]	[Φ] **/h/** *(futariwa, ofu)*	[ɧˣ, f, fᵠ, x, hʷ]

Table 1a. Comparison of the Swedish sounds in the imitators' Swedish repertoire which might be treated as similar to the Japanese target [ɸ], and the imitators' actual imitative productions.

Swedish has two non-strident fricatives: [ɧʷ] (a (usually rounded) velar-bidental fricative) and [f]. The similar Japanese sound [ɸ] was produced in a number of different ways by the subjects, ranging from a rounded glottal fricative to the (predicted) similar L1 sound [f] (Table 1a). All imitations shared one characteristic, namely non-stridency, but contrary to the SLM they did not use the Swedish 'equivalents' with the exception of a couple of instances of [f]. Instead, subjects created new (but

incorrect) sounds as attempts at the target sound, though often *based* on existing Swedish sounds. Perhaps most importantly, the subjects did not all hypothesise in the same way.

A second similar sound, usually realised as a type of tap [ɾ], would be predicted to be treated as equivalent to Swedish [r] or [d]. In this case, imitations were either correct, or consisted of a trill or voiced stop (Table 1b). An alveolar trill is possible in Swedish, but not used by any of these subjects in their respective dialects (they have either a uvular trill or approximant, or a retroflex fricative). Thus, the subjects again behaved predominantly contrary to the SLM's predictions.

predicted perception	*Swedish inventory*	***Japanese stimulus***	*imitators' output*
1b similar	[(r), d]	[ɾ] /ɾ/ *(futariwa, pairotto)*	[ɾ, r, d]

Table 1b. Comparison of the Swedish sounds in the imitators' Swedish repertoire which might be treated as similar to the Japanese target [ɾ], and the imitators' actual imitative productions.

The identical sound [ɕ] was produced either as the target, or as an English-like palatoalveolar (Table 2). That subjects seem to use other acquired L2 sounds in attempting a new sound is a phenomenon barely touched upon previously (cf Hammarberg & Hammarberg (1993)).

predicted perception	*Swedish inventory*	***Japanese stimulus***	*imitators' output*
2 identical	[ɕ]	[ɕ] /s/ *(hanashite, hashi)*	[ɕ, ʃ]

Table 2. Comparison of the Swedish sounds in the imitators' Swedish repertoire which might be treated as similar to the Japanese target [ɕ], and the imitators' actual imitative productions.

Further evidence for learners accessing both L1 and existing L2 sound representations in the decoding of new input is further supported by these subjects' behaviour in imitating the new sound [ɤ̆] (Table 3). This target was uniformly produced as a labiovelar approximant [w], indicating that learners classified the sound as similar rather than new, on the basis of its prior existence in their *accumulated* phonetic repertoire (which includes English).

predicted perception	*Swedish inventory*	***Japanese stimulus***	*imitators' output*
3 new	no comparable sound	[ɤ̆] /w/ *(wa)*	[w]

Table 3. Comparison of the Swedish sounds in the imitators' Swedish repertoire which might be treated as similar to the Japanese target [ɸ], and the imitators' actual imitative productions.

The data presented thus far are from speakers with no phonological knowledge of Japanese — they have no structural aids — and can be regarded as representative of learners at the absolute beginning of learning Japanese. Their productions show that beginner learners do not reliably conform to the predictions of Flege's Speech Learning Model in its present form, as they innovate where it is predicted that they will substitute an existing L1 sound, at times even where the L2 target is identical to an L1 sound, and they substitute familiar sounds from their entire phonetic repertoire for an L2 target which they are predicted to have no problems with.

But what of phonetic problems in more advanced learners? Certainly, they show behaviour in accordance with the SLM — for instance, Swedish speakers of English often use [s] for English [z], as it is the nearest available sound in their Swedish inventory. But they also diverge from the predicted behaviour, albeit less radically than the above imitators did for [ɸ], and can behave in what initially appear to be utterly unpredictable ways.

word examples	*Swedish inventory*	***English target***	*output*
cheese, chop, chuck	[ɕ, (tɕ)]	[ʧ]	[ɕ]

Table 4. Some Swedish learners' production of English /ʧ/, compared to the similar sounds present in these learners' Swedish repertoire.

Despite massive exposure to English, some Swedish speakers make errors on sounds represented entirely consistently in English orthography. The target affricate [tʃ] is mispronounced as the closest Swedish fricative [ɕ] (Table 4). This occurs despite the fact that all Swedish speakers are used to hearing an affricate [tɕ] instead of the aforementioned fricative in some older speakers' speech, in certain words, and in some dialects of Swedish (especially Finland-Swedish). Although this sound must at the very least be represented in their perceptual phonetic system, these poor speakers of English do not or cannot access the affricate for production. It is perhaps possible that this occurs because the difference between the Swedish affricate and fricative is not perceptually salient, as they are freely varying allophones of the same phoneme. This explanation is however problematic for the SLM, as the sound level addressed by the SLM is an (almost) allophonic one.

word examples	*Swedish inventory*	***English target***	*output*
young, year, you	[j̟, (j)]	[j̟]	[j̟, ʤ]

Table 5. Some Swedish learners' production of English /j/, compared to the similar sounds present in these learners' Swedish repertoire.

Another orthographically transparent sound in English which can present problems for some Swedish speakers is the voiced palatal approximant [j] (Table 5). In this case, most Swedes actually have an identical sound available, as a free variant of what is usually pronounced as a voiced palatal fricative [j̟]. Despite this, the speakers in question consistently use the fricative rather than the approximant when speaking English, or innovate and use an alveolopalatal affricate [ʤ] which is perceptually and productionally further away from the target approximant. This might be called *erroneous innovation*. The implication is that the speaker treats a sound which should be identical or similar as new, yet the likelihood that these speakers are actually hearing English [j] as an affricate is negligible, given that Swedish has no voiced affricates (with which it might then be confused). A parallel can be drawn between this example and that of the imitations of the 'identical' Japanese sound [ɕ]. Perhaps an L2 sound can seem 'too close' to the L1, and the learner chooses to *diverge* from the target rather than make what he suspects would be a probable error by using an existing L1 sound. I do not, however, find this explanation to be intuitively attractive.

word examples		***English target***	*output*
	Swedish inventory		
run, Gary, merry	[ɹ̟, ʒ̟, ʁ, r]	[ɹ]	[ʒ̟, ɹ, ʊ]
	German inventory		
	[ʁ]		[ʁ, ɹ, ʊ, w]

Table 6. Some Swedish and German learners' production of English /ɹ/, compared to the similar sounds present in these learners' Swedish and German repertoires.

The final example to be presented here relates to a phenomenon one might call *native-like innovation*. Some Swedish and German speakers of British English use a labial or labiodental approximant [ʊ] for English /r/ (Table 6). This is of course a relatively common allophone in native British English speakers, and one might assume that the non-native speakers have learnt this allophone from exposure. However, [ʊ] is strongly underrepresented in the British television programmes broadcast in Sweden, and American programmes predominate, whilst German learners are not usually exposed to English on television as almost everything is dubbed. Furthermore, the model provided in school will have been the more common (and prescriptively more approved of) postalveolar approximant [ɹ]. These speakers are thus producing a native allophone to which they will have had no or minimal exposure.

DISCUSSION

The data presented show that the Speech Learning Model does not provide a workable predictive construct. Although the ideas of newness and similarity are intuitively attractive, Flege's criteria for their application are too weak to account for readily observable learner behaviour. Given that simple L1-L2 categorial comparison is insufficient to account for these data—as also observed by Polka (1991) and Munro (1993)—we must seek explanations in other parts of phonetic and phonological representation and processing. As found by Polka, acoustic characteristics could be the strongest cues to any learner judgement of newness or similarity. This is supported both by my subjects' imitations of Japanese [ɸ], which consistently reflected the low-frequency frication (non-stridency) of

the target, and by the observation of [ʊ]-use by learners of British English where it can be hypothesised that learners show the same perception of [ɪ] as some native English speakers have made in acquiring their first language (where labiality is analysed as a salient characteristic).

On the basis of the data presented here, I would claim that learners use an analysis process consisting of at least the following elements:

1. Acoustic analysis

2. Comparison of analysis to *all* available 'sound categories' (unclear how close to surface reality)

3. Assessment of input *distance/proximity* from categories

If a model of sound learning takes into consideration the fact that learners often have experience of multiple dialects and/or other languages, then it should be easier to predict learner behaviour. Clear support for this is evident from the data shown in Tables 2 and 3. Given the fact that learners do not all produce the same sounds in attempting to pronounce a sound in another language, it can in part be concluded that these learners each have their own representational system. Although this statement is intuitively correct, both Flege's SLM and another similar model (Best's Perceptual Assimilation Model (Best, 1995) ignore this possibility. Furthermore, I would claim that the individual variation seen can be interpreted as favouring an input assessment based on input-to-category proximity measures, rather than the fairly simple binary decision implied by the SLM. This is further supported by the fact that *individual learners* varied in their multiple imitation attempts. A theory of perception based on categorial 'magnets' which attract new perceptual stimuli as a function of the magnet strength (weight) and the proximity of the stimulus has been proposed by Patricia Kuhl (cf Kuhl & Iverson, 1995), and would seem to be an attractive means to account for some of the learner behaviour described. This would also allow for the differences in learner hypothesisation about the input, as magnets may be weighted differently due to differences in a learner's total perceptual experience, and as Kuhl's Native Language Magnet theory appears to assume an acoustic analysis of what is perceived, learner differences can also be explained by ambiguities in the signal which each learner must resolve—such as may be the case for English /ɪ/ or Japanese [ɸ]. (See eg Ohala (1986) for a discussion of ambiguity.)

CONCLUSION

The data presented from learners of Japanese and English show that Flege's SLM is inadequate for predicting learner behaviour in learning 'foreign' sounds. The concepts of newness and similarity, whilst intuitively attractive, are not viable without empirical testing and identification of the perceptual information and criteria used by learners in analysing input. Some studies indicate that certain acoustic characteristics play a role in the classification of new sounds by learners, thereby showing that perceptual criteria for newness and similarity may be at least partially quantifiable. However, learner-specific factors are also evident in the data presented in this paper, indicating the need for better understanding of individual representation of phonetic information, and how representations interact, before predictions can be made about learner behaviour.

REFERENCES

Best, C.T. (1995): A direct-realist view of cross-language speech perception. In W. Strange, ed, *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues.* (pp 171-204). (Timonium, MD: York Press)

Flege, J.E. (1992): The intelligibility of English vowels spoken by British and Dutch talkers. In R.D. Kent, ed, *Intelligibility in Speech Disorders: Theory, Measurement, and Management.* (pp 157-232). (Philadelphia: John Benjamins)

Hammarberg, B. & Hammarberg, B. (1993): Articulatory re-setting in the acquisition of new languages. *PHONUM (Reports from the Department of Phonetics, University of Umeå),* **2**: 61-67.

Kuhl, P.K. & Iverson, P. (1995): Linguistic experience and the "perceptual magnet effect". In W. Strange, ed, *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues.* (pp 121-154). (Timonium, MD: York Press)

Markham, D. (1995): Investigating imitative ability. In K. Elenius & P. Branderud, eds, *The XIIth International Congress of Phonetic Sciences* (Vol.1, pp 314-317). (Stockholm: Stockholm University/KTH)

Munro, M.J. (1993): Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech,* **36**(1): 39-66.

Ohala, J.J. (1986): Against the direct realist view of speech perception. *Journal of Phonetics,* **14**(1): 75-82.

Polka, L. (1991): Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *JASA,* **89**(6): 2961-2977.