

# Negative transfer effect by cantonese students learning english



**ASSIGN  
BUSTER**

This work[1] is focused on deriving salient mispronunciations made by Cantonese (L1) learners of English (L2) specifically due to negative language transfer effect . The study presents phonological comparisons between Cantonese and American English along with the anecdotal examples observed in daily conversations.

In his study[2], Flege proposed a speech learning model for L2 pronunciation acquisition. The postulates and hypotheses forming this model are derived from empirical research concerning the differences in production and perception of sounds in L1 and L2. The model is based on the assumption that the phonetic systems responsible for production and perception of phonetic units, reorganize itself to incorporate new sounds through addition of new phonetic categories or through modification of the sound that exist in native language of an L2 speaker.

Strategies for providing meaningful articulatory feedback in the context of computer assisted pronunciation training are explored in this paper[3]. The study presents a survey of language teachers' and their students' experience in using articulatory instructions and feedback generated by a virtual tutor. When compared to a teacher delivering generalized instructions for a whole class, CAPT has better potentials to meet the individual's need by letting the learner decide the amount and type of feedback provided.

Apart from secondary language Learning, Articulatory feedback generation has applications in speech pathology diagnosis and treatment. It is suggested that place-of-articulation errors related articulatory feedback produced by Electromagnetic Articulography(EMA) can be used to treat

speech motor pathologies and oral apraxia in individuals with brain damage after stroke [4].

2D and 3D visual articulatory models can be effective tools for Pronunciation Training and treatment of Speech Disorders[5]. The ability of children while recognizing speech sounds on the basis of visual articulatory models was tested. It is concluded that the complex 3D display of the articulatory movements provides no significant advantage over a simple 2D-midsagittal displays of vocal tract articulatory movements because it comprises all the necessary visual information of articulatory movements with less complex information to process for children.

Reconstructed from multiple coronal cross-sectional slices of the tongue, 3D tongue surfaces for sustained vocalization of selected English sounds are presented in [6]. The study found that four classes of tongue shape - front raising, complete groove, back raising and two-point displacement, and three classes of tongue-palate contact - bilateral, cross-sectional and combination of two, were enough to categorize the sounds examined. The conclusion obtained was that the shape of the tongue and its position with respect to palate is distinctively different for consonants versus vowels in a limited number of ways.

Even though tongue is highly mobile and the tongue muscles are arranged in complex manner, its movement is controlled by small number of independent muscle groups. Utilizing this established hypothesis, a controlled model of human tongue movement in speech was investigated, based on finite elements modeling of soft tissues[7]. The model-based

<https://assignbuster.com/negative-transfer-effect-by-cantonese-students-learning-english/>

analysis of empirical data helped the authors to identify a number of independent tongue articulators and muscle synergies. It can be inferred from the study that the movements of tongue can be represented by a simple model, in which the muscle commands are obtained by central commands (produced by CNS) through a linear relationship, where the independent variables are values representing the position of the jaw and the hyoid bone; and the tongue posture.

In the direction of quantifying the articulatory distinctiveness, this study [8] presents the classification accuracies and distinctiveness distance metric for English vowels and consonants based on tongue and lip movement time series data. The articulatory vowel and consonants spaces were derived based on the pairwise distinctiveness distance, which can serve as an objective measure of the severity of articulatory impairment.

The primitive actions of the vocal tract articulators called “ gestures” are the basic unit of phonological structures. The gestures are the pre-linguistic discrete units of actions that are inherent in the maturation of a developing child. The details about these gestures can be read in these excellent overview [9] or [10]. The studies provide detailed formal characterization of gestures as phonological units within the context of computational model.

Representation of the speech event in terms of time-varying articulatory variables is discussed in [11] . With the assumption that the vowels are specified in terms of variables representing the positions of articulators and consonants can be implemented as transformations on the underlying vowel-derived articulatory states satisfying the given constraints, an intelligible

speech signal is synthesized using the proposed model. The model is also used for VCV (Vowel-Consonant-Vowel) speech production with the aid of rules for articulator movement and the articulator-position to vocal-tract cross-sectional-area transformation to demonstrate its use in co-articulation.

During Audio-visual speech understanding sessions, tongue is generally not visible. However, it is well established that it carries significant part of the articulatory information. This study [12] show that the human subjects are able to utilize the tongue shape vision for identification of phonemes, especially for the case when the audio signal is strongly degraded or absent.

Utilizing the gestural description of an utterance, the articulatory movements are generated from the proposed 3D vocal tract model as an integral part of the Articulatory speech synthesizer [13]. The model consists of 7 wireframe meshes representing the 3D surfaces of the articulatory space. The shape of these meshes is determined by 23 parameters. The authors also present a gestural dominance model to simulate the co-articulation effect. The 3D geometric vocal tract model and the gestural dominance model are used to generate the parameter curves which yield realistic visual representation of speech as well as natural sounding speech output.

The tongue is composed of several functionally independent articulatory regions. This functional regionality within the tongue has been examined by covariance-based analysis which quantified the strength of coupling among four different tongue locations [14]. Phonemic differentiation in vertical tongue movement was reflected in the coupling value across pellet pairs with place of articulation. The study also identified distinct clustered groups

of tongue displacements for speech and swallowing based on their coupling profiles.

Understanding the effectiveness of articulatory features in pronunciation analysis and training, researchers have sought to develop acoustic-to-articulatory feature inversion system[15] , which can be beneficial for both perception and production of speech related studies. The authors in [16] developed an HMM-based acoustic-to-articulatory inversion approach for a visual articulatory feedback system. The inversion system is jointly trained on acoustic and EMA articulatory data. The quality of the articulatory trajectories was evaluated by measuring the performance of an “ articulatory HMM-based phonetic decoder”.

The difference in articulatory movements in first and second language can be the basis for studies in mispronunciation errors by L2 speakers. An fMRI (functional Magnetic Resonance Imaging) based study in speech of L2 speakers was done in [17]. The research found that there is increased activity in the specific regions of brain-planum temporale and parietal operculum, when the speaker is speaking in L2 as compared to when he/she is speaking in L1. Production of L2 speech possibly involve extra linguistic or cognitive processing hence it is less automatic and erroneous.

In [18] the authors compare articulatory trajectories between three groups of speakers - native English, German and Dutch, speaking English. The study was carried out for a total of 69 speakers. Normalized tongue position difference trajectories were plotted for the sound pairs /t/-/θ/ and /s/-/ʃ/. The articulatory results reveal that for the Dutch speakers, there is no significant

difference between the sound pairs. Both ASR and human listener based perception results for confusion between the sounds of interest are presented. The perception results further show higher rate of misrecognition for the Dutch speakers as compared to English and German speakers.

In a work closer to the objective of our paper, EMA Articulatory data is compared between a Mandarin speaker of English and a native speaker of English in [19]. For the English phonemes not present in Mandarin phonemic inventory, pairwise Mahalanobis distance between the displacements of the articulators on tongue (3 points) and lips (3 points) is calculated between the two speakers. The dissimilarity information visualized from Hierarchical clustering analysis (HCA) and multi-dimensional scaling (MDS) clearly show significant difference in articulation between the Mandarin sounds and their English equivalents.

A 3D talking head system built from facial video and X-ray database capturing the movement of external and internal articulators respectively was proposed in [20]. Key images reflecting the significant differences in articulator positions among selected English sounds were chosen to build a 3D model. A set of commonly confused phonemes for Chinese speakers of English were selected to evaluate the 3D talking head model. The evaluation is done by the subjects who were asked to identify the word for the displayed animation and score its realistic degree. The word recognition accuracy by the subjects for the proposed 3D talking head model was 77.3%. The implications of this work can be in the direction of generation of realistic articulatory feedback for Computer Aided Pronunciation Training systems.

With relatively larger sample size of 34 speakers used for Articulography study, the authors in [21] analyze articulatory differences in two different Dutch dialects. Curve fitted to the data points for tongue trajectories for two groups of speakers reveal clear distinction between the dialects.

Analysis of mutual influence between L1 (Mandarin) and L2 (English) phonetic system in bilingual children was carried out in [22]. Static and dynamic spectral features were taken into consideration for the study. It can be inferred from the results that bilingual children tend to carry their L1 features towards L2 in the beginning phase of L2 pronunciation learning. However, when those children become highly proficient in L2, they not only are able to produce sounds in native-like manner, but also tend to transfer L2 features to their L1.

Phoneme-level articulator dynamics serve as fundamental information in 3D articulatory system. Therefore, in order to design an animation model in pronunciation training, a HMM-based model is proposed in [23, 24]. The proposed system in [23] takes speech as input and generates articulatory movement trajectories at phoneme-level. These curves are transformed to points in the 3D articulatory animation model. Comparison of synthesized articulatory contours versus the EMA generated curve for frequently confused phoneme pairs in Mandarin (L1)-English (L2) demonstrate the capability of the system to synthesize phoneme-level articulator dynamics used in transparent talking head animation for pronunciation training.

In an excellent tutorial on articulatory differences between native and non-native speakers of English using dynamic phonetic data, the author

<https://assignbuster.com/negative-transfer-effect-by-cantonese-students-learning-english/>



describes about a generalized additive mixed modeling (GAM) approach for analysis of articulatory data[25] . GAM is a regression model with capability to capture non-linear patterns in data. The tutorial gives hands-on description of how to use GAM modeling in the R package, *mgcv* .

1. Meng, H., E. Zee, and W. S. Lee, *A contrastive phonetic study between Cantonese and English to predict salient mispronunciations by Cantonese learners of English*. Unpublished article. The Chinese University of Hong Kong, 2007.
2. Flege, J. E., *Second language speech learning: Theory, findings, and problems*. Speech perception and linguistic experience: Issues in cross-language research, 1995. 92: p. 233-277.
3. Engwall, O., *Feedback strategies of human and virtual tutors in pronunciation training*. Speech Music and Hearing---Quarterly Progress and Status Report, 2006. 48(1).
4. Katz, W. F. and M. R. McNeil, *Studies of articulatory feedback treatment for apraxia of speech based on electromagnetic articulography*. Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders, 2010. 20(3): p. 73-79.
5. Kröger, B. J., V. Graf-Borttscheller, and A. Lowit. *Two and three-dimensional visual articulatory models for pronunciation training and for treatment of speech disorders* . in *Interspeech, 9th Annual Conference of the International Speech Communication Association* . 2008.

6. Stone, M. and A. Lundberg, *Three-dimensional tongue surface shapes of English consonants and vowels*. The Journal of the Acoustical Society of America, 1996. 99(6): p. 3728-3737.
7. Sanguineti, V., R. Laboissiere, and Y. Payan, *A control model of human tongue movements in speech*. Biological cybernetics, 1997. 77(1): p. 11-22.
8. Wang, J., et al., *Articulatory distinctiveness of vowels and consonants: A data-driven approach*. Journal of Speech, Language, and Hearing Research, 2013.
9. Browman, C. P. and L. Goldstein, *Articulatory gestures as phonological units*. Phonology, 1989. 6(2): p. 201-251.
10. Browman, C. P. and L. Goldstein, *Articulatory phonology: An overview*. Phonetica, 1992. 49(3-4): p. 155-180.
11. Mermelstein, P., *Articulatory model for the study of speech production*. The Journal of the Acoustical Society of America, 1973. 53(4): p. 1070-1082.
12. Badin, P., et al., *Can you 'read'tongue movements? Evaluation of the contribution of tongue display to speech understanding*. Speech Communication, 2010. 52(6): p. 493-503.
13. Birkholz, P., D. Jackèl, and B. J. Kroger. *Construction and control of a three-dimensional vocal tract model* . in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings* . 2006. IEEE.

14. Green, J. R. and Y.-T. Wang, *Tongue-surface movement patterns during speech and swallowing*. The Journal of the Acoustical Society of America, 2003. 113(5): p. 2820-2833.
15. Engwall, O. *Pronunciation analysis by acoustic-to-articulatory feature inversion* . in *International Symposium on Automatic Detection of Errors in Pronunciation Training* . 2012.
16. Youssef, A. B., et al. *Toward a multi-speaker visual articulatory feedback system* . 2011.
17. Simmonds, A. J., et al., *A comparison of sensory-motor activity during speech in first and second languages*. Journal of neurophysiology, 2011. 106(1): p. 470-478.
18. Wieling, M., et al. *Articulatory differences between L1 and L2 speakers of English* . in *Proceedings of The 11th International Seminar on Speech Production, Tianjin, China, October* . 2017.
19. Li, S. and L. Wang. *Cross linguistic comparison of Mandarin and English EMA articulatory data* . in *Thirteenth Annual Conference of the International Speech Communication Association* . 2012.
20. Wang, L., H. Chen, and J. Ouyang. *Evaluation of external and internal articulator dynamics for pronunciation learning* . in *Tenth Annual Conference of the International Speech Communication Association* . 2009.
21. Wieling, M., et al., *Investigating dialectal differences using articulography*. Journal of Phonetics, 2016. 59: p. 122-143.

22. Yang, J. and R. A. Fox, *L1-L2 interactions of vowel systems in young bilingual Mandarin-English children*. *Journal of Phonetics*, 2017. 65: p. 60-76.
23. Wang, L., et al., *Phoneme-level articulatory animation in pronunciation training*. *Speech Communication*, 2012. 54(7): p. 845-856.
24. Li, S., L. Wang, and E. Qi. *The phoneme-level articulator dynamics for pronunciation animation* . in *2011 International Conference on Asian Language Processing* . 2011. IEEE.
25. Wieling, M., *Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English*. *Journal of Phonetics*, 2018. 70: p. 86-116.