Bibliography

[1] Abe, M., Nakamura, S., Shikano, K., and Kuwabara, H. Voice conversion through vector quantization. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing(ICASSP)*, pages 655–658. IEEE, 1988.

[2] Aihara, R., Takashima, R., Takiguchi, T., and Ariki, Y. GMM-based emotional voice conversion using spectrum and prosody features. *American Journal of Signal Processing*, 2(5):134–138, 2012.

[3] Al-Tamimi, J. Static and dynamic cues in vowel production: A cross dialectal study in Jordanian and Moroccan Arabic. In *16th International Congress of Phonetic Sciences (ICPhS), Saarbrücken, Germany. Paper retrieved from www. ddl. ish-lyon. cnrs. fr/fulltext/al.../al-tamimi_2007_icphs_draft. pdf*, 2007.

[4] Al-Tamimi, J. Spectral tilt as an acoustic correlate to pharyngealisation in Jordanian and Moroccan Arabic. 2015.

[5] AlDahri, S. S. A study for the effect of the Emphaticness and language and dialect for Voice Onset Time (VOT) in Modern Standard Arabic (MSA). *arXiv preprint arXiv:1305.2680*, 2013.

[6] Allen, J., Hunnicutt, M. S., Klatt, D. H., Armstrong, R. C., and Pisoni, D. B. *From text to speech: The MITalk system*. Cambridge University Press, 1987.

[7] Anumanchipalli, G. K., Oliveira, L. C., and Black, A. W. A style capturing approach to F0 transformation in voice conversion. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6915–6919. IEEE, 2013.

[8] Aryal, S. and Gutierrez-Osuna, R. Can voice conversion be used to reduce non-native accents? In *Proceedings of the IEEE International Conference on*

Acoustics, Speech and Signal Processing (ICASSP), pages 7879–7883. IEEE, 2014.

[9] Bahmaninezhad, F., Sameti, H., and Khorram, S. Hmm-based Persian speech synthesis using limited adaptation data. In *Proceedings of the IEEE 11th International Conference on Signal Processing (ICSP)*, volume 1, pages 585–589. IEEE, 2012.

[10] Barreda, S. Investigating the use of formant frequencies in listener judgments of speaker size. *Journal of Phonetics*, 55:1–18, 2016.

[11] Benkı, J. R. Place of articulation and first formant transition pattern both affect perception of voicing in English. *Journal of Phonetics*, 29(1):1–22, 2001.

[12] Beskow, J. and Gustafson, J. Experiments with Synthesis of Swedish Dialects. *FONETIK, Stockholm*, pages 28–29, 2009.

[13] Black, A. W., Bunnell, H. T., Dou, Y., Perry, D., Polzehl, T., et al. New parameterizations for Emotional Speech Synthesis. 2011.

[14] Bob, M. Klattworks: A [somewhat] new systematic approach to formant-based speech synthesis for empirical research. 2009.

[15] Boersma, P. and Weenink, D. Praat: doing phonetics by computer. 2013. *Available form: URL: http://www. fon. hum. uva. nl/praat*, 2013.

[16] Bõhm, T. and Németh, G. Algorithm for formant tracking, modification and synthesis. *GYULA SALLAI–president, Scientific Association for Infocommunications ÁKOS DETREKÔI–president, National Council of Hungary for Information and Communications Technology*, page 15, 2007.

[17] Bollepalli, B., Beskow, J., and Gustafson, J. Non-linear Pitch Modification in Voice Conversion Using Artificial Neural Networks. In *Advances in Nonlinear Speech Processing*, pages 97–103. Springer, 2013.

[18] Brown, L. and Yeon, J. *The handbook of Korean linguistics*. John Wiley & Sons, 2015.

[19] Bulyko, I., Ostendorf, M., and Price, P. On the relative importance of different prosodic factors for improving speech synthesis. In *Proceedings of ICPhs*, volume 99, page 81, 1999.

[20] Burkhardt, F. and Sendlmeier, W. F. Verification of acoustical correlates of emotional speech using formant-synthesis. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, 2000.

[21] Chao-angthong, P., Suchato, A., and Punyabukkana, P. Northern Thai Dialect Text to Speech. In *Proceedings of the 14th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pages 1–6. IEEE, 2017.

[22] Chen, G. *The Voice Source in Speech Production: From Models to Applications*. University of California, Los Angeles, 2014.

[23] Chen, G., Garellek, M., Kreiman, J., Gerratt, B. R., and Alwan, A. A perceptually and physiologically motivated voice source model. In *Proceedings of Meetings on Acoustics ICA2013*, volume 19, page 060160. ASA, 2013.

[24] Chen, L.-m., Chao, K.-Y., and Peng, J.-F. VOT productions of word-initial stops in Mandarin and English: A cross-language study. In *ROCLING*, 2007.

[25] Chiang, C.-Y. Cross-Dialect Adaptation Framework for Constructing Prosodic Models for Chinese Dialect Text-to-Speech Systems. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2017.

[26] Childers, D. G., Wu, K., Hicks, D., and Yegnanarayana, B. Voice conversion. *Speech Communication*, 8(2):147–158, 1989.

[27] Cho, T. and Ladefoged, P. Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, 27(2):207–229, 1999.

[28] Clopper, C. G., Pisoni, D. B., and De Jong, K. Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America*, 118(3):1661–1676, 2005.

[29] Collomb, C. Burg's method, algorithm and recursion. *K. Elissa, unpublished*, 2010.

[30] Conway, E. A. An analysis of Somali pronunciation errors. 2008.

[31] Desai, S., Black, A. W., Yegnanarayana, B., and Prahallad, K. Spectral mapping using artificial neural networks for voice conversion. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(5):954–964, 2010.

[32] Donovan, R. E. *Trainable speech synthesis*. PhD thesis, University of Cambridge, 1996.

[33] Dorn, A. and Chasaide, A. N. Effects of Focus on f_0 and Duration in Irish (Gaelic) Declaratives. In *Proceedings of Twelfth Annual Conference of the International Speech Communication Association*, 2011.

[34] Enflo, L. Spectral Tilt used for Automatic Detection of Prominence: A Comparison between Electroglottography and Audio. *Term Paper for Machine Learning, a GSLT course*, 2009.

[35] Ewald, O., Asu, E. L., and Schötz, S. The formant dynamics of long close vowels in three varieties of Swedish. In *Interspeech 2017*, pages 1412–1416, 2017.

[36] Fant, G. *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations*. Mouton & Company, 1970.

[37] Ferragne, E. and Pellegrino, F. Diphthongization as cue for the automatic identification of British English dialects. *The Journal of the Acoustical Society of America*, 116(4):2630–2630, 2004.

[38] Flynn, N. and Foulkes, P. Comparing vowel formant normalization methods. In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, pages 683–686. City University of Hong Kong) (Hong Kong, 2011.

[39] Fox, R. A. and Jacewicz, E. Analysis of total vowel space areas in three regional dialects of American English. *Journal of the Acoustical Society of America*, 123(5):3068, 2008.

[40] Fox, R. A. and Jacewicz, E. Cross-dialectal variation in formant dynamics of American English vowels. *The Journal of the Acoustical Society of America*, 126(5):2603–2618, 2009.

[41] Frarrar, K., Grabe, E., and Nolan, F. English intonation in the British isles. *Leeds Studies in English*, 30:243, 1999.

[42] Godoy, E., Rosec, O., and Chonavel, T. Voice conversion using dynamic frequency warping with amplitude scaling, for parallel or nonparallel corpora. *IEEE Transactions on Audio, Speech and Language Processing*, 20(4):1313–1323, 2012.

[43] Gopinath, D. P., Veena, S., and Nair, A. S. Modeling of Vowel Duration in Malayalam Speech using Probability Distribution. *ISCA Archive, Speech Prosody, Campinas, Brazil*, pages 6–9, 2008.

[44] Gordon, M. and Ladefoged, P. Phonation types: a cross-linguistic overview. *Journal of phonetics*, 29(4):383–406, 2001.

[45] Goswami, G. C. and Tamuli, J. Asamiya. In Cardona, G. and Jain, D., editors, *The Indo-Aryan Languages*, pages 391–443. Routledge, London, 2003.

[46] Groner, G., Bernstein, J., Ingber, E., Pearlman, J., and Toal, T. A real-time text-to-speech converter. *Speech Technology*, 1(2):73–76, 1982.

[47] Gu, C. et al. Smoothing spline anova models: R package gss. *Journal of Statistical Software*, 58(5):1–25, 2014.

[48] Hagiwara, R. Dialect variation and formant frequency: The American English vowels revisited. *The Journal of the Acoustical Society of America*, 102(1): 655–658, 1997.

[49] Hallahan, W. I. DECtalk software: Text-to-speech technology and implementation. *Digital Technical Journal*, 7(4):5–19, 1995.

[50] Hart, J. *Prosodic differences among dialects of American English*. PhD thesis, The Ohio State University, 2013.

[51] Hershey, J. R. and Olsen, P. A. Approximating the Kullback Leibler divergence between Gaussian mixture models. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 4, pages IV–317. IEEE, 2007.

[52] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N., et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, 29(6):82–97, 2012.

[53] Hojo, N., Ijima, Y., and Mizuno, H. An Investigation of DNN-Based Speech Synthesis Using Speaker Codes. In *INTERSPEECH*, pages 2278–2282, 2016.

[54] Holliday, J. J. and Kong, E. J. Dialectal variation in the acoustic correlates of Korean stops. In *Proceedings of the 17th International Congress of Phonetic Sciences*, pages 878–881, 2011.

[55] Holt, Y., Jacewicz, E., and Fox, R. Temporal variation in African American English: The Distinctive Use of Vowel Duration. *J Phonet Audiol*, 2(121):2, 2016.

[56] Hou, J., Liu, Y., Zheng, T. F., Olsen, J., and Tian, J. Using cepstral and prosodic features for Chinese accent identification. In *Proceedings of the 7th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, pages 177–181. IEEE, 2010.

[57] Hung, P. N., Loan, T. V., and Quang, N. H. Automatic identification of Vietnamese dialects. *Journal of Computer Science and Cybernetics*, 32(1): 19–30, 2016.

[58] Hunt, A. J. and Black, A. W. Unit selection in a concatenative speech synthesis system using a large speech database. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 373–376. IEEE, 1996.

[59] Hwang, H.-T., Tsao, Y., Wang, H.-M., Wang, Y.-R., and Chen, S.-H. Incorporating global variance in the training phase of GMM-based voice conversion. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific*, pages 1–6. IEEE, 2013.

[60] Inanoglu, Z. Transforming pitch in a voice conversion framework. *M.Phil Dissertation, St. Edmonds College, University of Cambridge*, 2003.

[61] Jacewicz, E. and Fox, R. A. Cross-dialectal differences in dynamic formant patterns in American English vowels. In *Vowel inherent spectral change*, pages 177–198. Springer, 2013.

[62] Jacewicz, E., Fox, R., and Salmons, J. Vowel duration in three American English dialects. *American Speech*, 82(4):367–385, 2007.

[63] Jacewicz, E., Fox, R. A., and Wei, L. Between-speaker and within-speaker variation in speech tempo of American English. *The Journal of the Acoustical Society of America*, 128(2):839–850, 2010.

[64] Jiao, Y., Xie, X., Na, X., and Tu, M. Improving voice quality of HMM-based speech synthesis using voice conversion method. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7914–7918. IEEE, 2014.

[65] Joo, Y.-S., Jun, W.-S., and Kang, H.-G. Efficient deep neural networks for speech synthesis using bottleneck features. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2016 Asia-Pacific*, pages 1–4. IEEE, 2016.

[66] Kain, A. and Macon, M. W. Spectral voice conversion for text-to-speech synthesis. In *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 1, pages 285–288. IEEE, 1998.

[67] Kakati, B. *Assamese, its formation and development.* Guwahati: LBS Publication, 1941.

[68] Kakouros, S., Räsänen, O., and Alku, P. Evaluation of Spectral Tilt Measures for Sentence Prominence Under Different Noise Conditions. *Proceedings of Interspeech 2017*, pages 3211–3215, 2017.

[69] Kang, S., Qian, X., and Meng, H. Multi-distribution deep belief network for speech synthesis. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 8012–8016. IEEE, 2013.

[70] Karjalainen, M., Altosaar, T., and Vainio, M. Speech synthesis using warped linear prediction and neural networks. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 2, pages 877–880. IEEE, 1998.

[71] Keerio, A., Dhomeja, L., Shaikh, A., and Malkani, Y. Comparative Analysis of Vowels, Diphthongs and Glides in Sindhi. *Signal and Image Processing : An International Journal (SIPIJ)*, 2(4), 2011.

[72] Kelly Jr, J. L. and Gerstman, L. J. An artificial talker driven from a phonetic input. *The Journal of the Acoustical Society of America*, 33(6):835–835, 1961.

[73] Khaw, Y.-M. J. and Tan, T.-P. A Fast Adaptation Technique for building Dialectal Malay Speech Synthesis Acoustic Model. *Jurnal Teknologi, Malaysia*, pages 57–66, 2015.

[74] Kiruthiga, S. and Krishnamoorthy, K. Annotating Speech Corpus for Prosody Modeling in Indian Language Text to Speech Systems. 2012.

[75] Klatt, D. The Klattalk text-to-speech conversion system. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'82.*, volume 7, pages 1589–1592. IEEE, 1982.

[76] Klatt, D. H. Software for a cascade/parallel formant synthesizer. *the Journal of the Acoustical Society of America*, 67(3):971–995, 1980.

[77] Koneva, E. V. and Gural, S. K. The role of dialects in the German society. *Procedia-Social and Behavioral Sciences*, 200:248–252, 2015.

[78] Konno, H., Kudo, M., Imai, H., and Sugimoto, M. Whisper to normal speech conversion using pitch estimated from spectrum. *Speech Communication*, 83: 10–20, 2016.

[79] Koolagudi, S. G., Vishwanath, B. K., Akshatha, M., and Murthy, Y. V. Performance Analysis of LPC and MFCC Features in Voice Conversion Using Artificial Neural Networks. In *Proceedings of the International Conference on Data Engineering and Communication Technology*, pages 275–280. Springer, 2017.

[80] Labov, W. *Principles of linguistic change, cognitive and cultural factors*, volume 3. John Wiley & Sons, 2011.

[81] Laskar, R. H., Chakrabarty, D., Talukdar, F. A., Rao, K. S., and Banerjee, K. Comparing ANN and GMM in a voice conversion framework. *Applied Soft Computing*, 12(11):3332–3342, 2012.

[82] Lazaridis, A., Zervas, P., and Kokkinakis, G. Segmental duration modeling for Greek speech synthesis. In *Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence,ICTAI 2007.*, volume 2, pages 518–521. IEEE, 2007.

[83] Lei, M., Yamagishi, J., Richmond, K., Ling, Z.-H., King, S., and Dai, L.-R. Formant-Controlled HMM-Based Speech Synthesis. In *INTERSPEECH*, pages 2777–2780, 2011.

[84] Lein, T., Kupisch, T., and van de Weijer, J. Voice onset time and global foreign accent in German–French simultaneous bilinguals during adulthood. *International Journal of Bilingualism*, 20(6):732–749, 2016.

[85] Lengerisa, A., Kainadaa, E., Baltazanib, M., and Iversonc, P. Dialectal effects on the perception of Greek vowels. 2015.

[86] Liao, W.-W. and Shen, J.-L. Improved Prosody Module in a Text-to-Speech System. In *ROCLING*, 2004.

[87] Lisker, L. and Abramson, A. S. A cross-language study of Voicing in Initial Stops: Acoustical Measurements. In *Proceedings of the 6th international congress of phonetic sciences*, volume 20, pages 384–422, 1964.

[88] MacKay, I. R. *Phonetics: The science of speech production*. Little, Brown, 1987.

[89] Magen, H. S. Perception of diphthongized vowels in Rhode Island English. In *Proceedings of 15th International Congress of Phonetic Sciences. Causal Productions: Barcelona (1453-1456)*, 2003.

[90] Mannepalli, K., Sastry, P. N., and Suman, M. MFCC-GMM based accent recognition system for Telugu speech signals. *International Journal of Speech Technology*, 19(1):87–93, 2016.

[91] Martony, J. and Franzen, O. Dept. for Speech, Music and Hearing. *STL-QPSR*, 7(1):017–022, 1966.

[92] Mayr, R. and Davies, H. A cross-dialectal acoustic study of the monophthongs and diphthongs of Welsh. *Journal of the International Phonetic Association*, pages 01–25, 2011.

[93] Miller, J. L. and Volaitis, L. E. Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, 46(6):505–512, 1989.

[94] Mitra, V., Sivaraman, G., Nam, H., Espy-Wilson, C., and Saltzman, E. Articulatory features from deep neural networks and their role in speech recognition. In *Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3017–3021. IEEE, 2014.

[95] Mizuno, H., Abe, M., and Hirokawa, T. Waveform-based speech synthesis approach with a formant frequency modification. In *Proceedings of 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1993. ICASSP-93.*, volume 2, pages 195–198. IEEE, 1993.

[96] Mohammadi, S. H. and Kain, A. An overview of voice conversion systems. *Speech Communication*, 2017.

[97] Narendranath, M., Murthy, H. A., Rajendran, S., and Yegnanarayana, B. Voice conversion using artificial neural networks. In *Automatic Speaker Recognition, Identification and Verification*, 1994.

[98] Nirmal, J., Patnaik, S., and Zaveri, M. A. Voice transformation using radial basis function. In *Proceedings of the Third International Conference on Trends in Information, Telecommunication and Computing*, pages 345–351. Springer, 2013.

[99] Nordhoff, S. *A grammar of Upcountry Sri Lanka Malay*. LOT, 2009.

[100] Okati, F., Helgason, P., and Jahani, C. Diphthongization in five Iranian Balochi dialects. *Orientalia Suecana*, 61:107–119, 2013.

[101] Olive, J. P. A new algorithm for a concatenative speech synthesis system using an augmented acoustic inventory of speech sounds. In *The ESCA Workshop on Speech Synthesis*, 1991.

[102] Oliveira, L. C. Estimation of source parameters by frequency analysis. In *EUROSPEECH*, 1993.

[103] Onaolapo, J., Idachaba, F., Badejo, J., Odu, T., and Adu, O. A simplified overview of text-to-speech synthesis. 2014.

[104] Paulo, S. and Oliveira, L. C. Automatic phonetic alignment and its confidence measures. In *Advances in Natural Language Processing*, pages 36–44. Springer, 2004.

[105] Peng, J.-F., Chen, L.-m., and Lee, C.-C. Tonal Effects on Voice Onset Time. 14(4):341–361, 2009.

[106] Pucher, M., Neubarth, F., Strom, V., Moosmuller, S., Hofer, G., Kranzler, C., Schuchmann, G., and Schabus, D. Resources for speech synthesis of Viennese varieties. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC'2010*, 2010.

[107] Pucher, M., Kerschhofer-Puhalo, N., and Schabus, D. Phone set selection for HMM-based dialect speech synthesis. In *Proceedings of the First Workshop on Algorithms and Resources for Modelling of Dialects and Language Varieties*, pages 65–69. Association for Computational Linguistics, 2011.

[108] Pucher, M., Kerschhofer-Puhalo, N., Schabus, D., Moosmüller, S., and Hofer, G. Language resources for the adaptive speech synthesis of dialects. *Proceedings of SIDG, Vienna, Austria*, 2012.

[109] Pucher, M., Xhafa, V., Dika, A., and Toman, M. Adaptive speech synthesis of Albanian dialects. In *Proceedings of the International Conference on Text, Speech, and Dialogue*, pages 158–164. Springer, 2015.

[110] Pucher, M., Lozo, C., and Moosmüller, S. Phone Mapping and Prosodic Transfer in Speech Synthesis of similar Dialect Pairs. In *28. Konferenz Elektronische Sprachsignalverarbeitung 2017, Saarbrücken*, pages 180 –185, 2017.

[111] Qian, Y., Fan, Y., Hu, W., and Soong, F. K. On the training aspects of deep neural network (dnn) for parametric tts synthesis. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 3829–3833. IEEE, 2014.

[112] Qutaish, F. Pharyngealization Spread in Ibbi Dialect of Yemeni Arabic: An Acoustic Study. *International Journal of Humanities and Social Sciences*, 3(9):2098, 2016. URL `http://waset.org/abstracts/Humanities-and-Social-Sciences`.

[113] Raitio, T. et al. Voice source modelling techniques for statistical parametric speech synthesis. 2015.

[114] Rami, M. K., Kalinowski, J., Stuart, A., and Rastatter, M. P. Voice onset times and burst frequencies of four velar stop consonants in Gujarati. *The Journal of the Acoustical Society of America*, 106(6):3736–3738, 1999.

[115] Rao, K. S. Voice conversion by mapping the speaker-specific features using pitch synchronous approach. *Computer Speech & Language*, 24(3):474–494, 2010.

[116] Rao, K. S. *Predicting Prosody from Text for Text-to-Speech Synthesis*. Springer Science & Business Media, 2012.

[117] Rao, K. S., Nandy, S., and Koolagudi, S. G. Identification of Hindi dialects using speech. *WMSCI-2010*, 2010.

[118] Reddy, V. R. and Rao, K. S. Intensity Modeling for Syllable Based Text-to-Speech Synthesis. In *Contemporary Computing*, pages 106–117. Springer, 2012.

[119] Reddy, V. R. and Rao, K. S. Prosody modeling for syllable based text-to-speech synthesis using feedforward neural networks. *Neurocomputing*, 171: 1323–1334, 2016.

[120] Rentzos, D., Vaseghi, S., Yan, Q., and Ho, C.-H. Parametric formant modelling and transformation in voice conversion. *International Journal of Speech Technology*, 8(3):227–245, 2005.

[121] Rickford, J. R. How linguists approach the study of language and dialect. *Retrieved on January*, 20:2013, 2002.

[122] Rothenberg, M. Voice onset time versus articulatory modeling for stop consonants. *Logopedics Phoniatrics Vocology*, 34(4):171–180, 2009.

[123] Sagisaka, Y. ATR v-talk speech synthesis system. *Proceedings of ICSLP, 1992*, 1992.

[124] Sandoval, S., Berisha, V., Utianski, R. L., Liss, J. M., and Spanias, A. Automatic assessment of vowel space area. *The Journal of the Acoustical Society of America*, 134(5):EL477–EL483, 2013.

[125] Sarwar, A., Ahmed, S., and Tarar, A. Diphthongs in Urdu language and analysis of their acoustic properties. Annual Student Report,CRULP:09–15, 2003.

[126] Shaikh, H., Mahar, J. A., and Malah, G. A. Digital Investigation of Accent Variation in Sindhi Dialects. *Indian Journal of Science and Technology*, 6 (10):5429–5433, 2013.

[127] Sharma, C., Mahar, and Talukdar, P. Dialect variation in Boro Language and Grapheme-to-Phoneme conversion rules to handle lexical lookup fails in Boro TTS System. *International Journal of Scientific and Research Publications*, 2(9), 2012.

[128] Shikano, K., Nakamura, S., and Abe, M. Speaker adaptation and voice conversion by codebook mapping. In *Circuits and Systems, 1991., IEEE International Sympoisum on*, pages 594–597. IEEE, 1991.

[129] Shue, Y.-L. and Alwan, A. A new voice source model based on high-speed imaging and its application to voice source estimation. In *Proceedings of the 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 5134–5137. IEEE, 2010.

[130] Shue, Y.-L., Kreiman, J., and Alwan, A. A novel codebook search technique for estimating the open quotient. In *Interspeech*, pages 2895–2898, 2009.

[131] Sinha, S., Agrawal, S. S., and Jain, A. Dialectal influences on acoustic duration of Hindi phonemes. In *Oriental COCOSDA held jointly with 2013 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE), 2013 International Conference*, pages 1–5. IEEE, 2013.

[132] Sinha, S., Jain, A., and Agrawal, S. S. Speech processing for Hindi dialect recognition. In *Advances in Signal Processing and Intelligent Recognition Systems*, pages 161–169. Springer, 2014.

[133] Sitaram, S. *Pronunciation modeling for synthesis of low resource languages.* PhD thesis, Ph. D. dissertation, Carnegie Mellon University, 2015.

[134] Srikanth, R., Bajibabu, B., and Prahallad, K. Duration modelling in voice conversion using artificial neural networks. In *Proceedings of the 19th International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 556–559. IEEE, 2012.

[135] Stylianou, Y., Cappé, O., and Moulines, E. Continuous probabilistic transform for voice conversion. *IEEE Transactions on Speech and Audio Processing*, 6(2):131–142, 1998.

[136] Suendermann, D., Höge, H., and Black, A. Challenges in Speech Synthesis. In *Speech Technology*, pages 19–32. Springer, 2010.

[137] Sunija, A., Rajisha, T., and Riyas, K. Comparative Study of Different Classifiers for Malayalam Dialect Recognition System. *Procedia Technology*, 24: 1080–1088, 2016.

[138] Takeda, K., Abe, K., and Sagisaka, Y. *On the basic scheme and algorithms in non-uniform unit speech synthesis.* Amsterdam, The Netherlands: North-Holland, 1992.

[139] Taylor, P. *Text-to-speech synthesis.* Cambridge university press, 2009.

[140] Teutenberg, J. and Watson, C. Vowel quality in accent modification. In *Proceedings of the Australian International Conference on Speech Science & Technology, University of Auckland, New Zealand*, pages 292–295, 2006.

[141] Toda, T., Ohtani, Y., and Shikano, K. Eigenvoice Conversion Based on Gaussian Mixture Model. 2006.

[142] Toda, T., Black, A. W., and Tokuda, K. Voice conversion based on maximum-likelihood estimation of spectral parameter trajectory. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(8):2222–2235, 2007.

[143] Türk, O. and Schröder, M. A comparison of voice conversion methods for transforming voice quality in emotional speech synthesis. In *INTERSPEECH 2008*, pages 2282–2285, 2008.

[144] Turk, O. and Schroder, M. Evaluation of expressive speech synthesis with voice conversion and copy resynthesis techniques. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(5):965–973, 2010.

[145] Uslu, I. B., Ilk, H. G., and Yilmaz, A. E. A rule based prosody model for Turkish text-to-speech synthesis. *Tehnicki Vjesnik*, 20(1):217–223, 2013.

[146] Vainio, M., Airas, M., Järvikivi, J., and Alku, P. Laryngeal voice quality in the expression of focus. In *Proceedings of Eleventh Annual Conference of the International Speech Communication Association*, 2010.

[147] Valbret, H., Moulines, E., and Tubach, J.-P. Voice transformation using PSOLA technique. *Speech communication*, 11(2-3):175–187, 1992.

[148] Violante, L., Zivic, P. R., and Gravano, A. Improving speech synthesis quality by reducing pitch peaks in the source recordings. In *HLT-NAACL*, pages 502–506, 2013.

[149] Wang, M., Wen, M., Hirose, K., and Minematsu, N. Emotional voice conversion for Mandarin using tone nucleus model–small corpus and high efficiency. In *Speech Prosody 2012*, 2012.

[150] Wang, Z., Wang, R., Shuang, Z., and Ling, Z. A novel voice conversion system based on codebook mapping with phoneme-tied weighting. In *Eighth International Conference on Spoken Language Processing*, 2004.

[151] Watts, O., Yamagishi, J., King, S., and Berkling, K. Synthesis of child speech with HMM adaptation and voice conversion. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(5):1005–1016, 2010.

[152] Wayland, R. and Jongman, A. Acoustic correlates of breathy and clear vowels: the case of Khmer. *Journal of Phonetics*, 31(2):181–201, 2003.

[153] Williams, D. and Escudero, P. A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *The Journal of the acoustical society of America*, 136(5):2751–2761, 2014.

[154] Wu, Z. and Li, H. Voice conversion and spoofing attack on speaker verification systems. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific*, pages 1–9. IEEE, 2013.

[155] Wu, Z., Virtanen, T., Chng, E. S., and Li, H. Exemplar-based sparse representation with residual compensation for voice conversion. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 22(10): 1506–1521, 2014.

[156] Wu, Z., Swietojanski, P., Veaux, C., Renals, S., and King, S. A study of speaker adaptation for DNN-based speech synthesis. In *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[157] Xydas, G., Spiliotopoulos, D., and Kouroupetroglou, G. Modeling improved prosody generation from high-level linguistically annotated corpora. *IEICE TRANSACTIONS on Information and Systems*, 88(3):510–518, 2005.

[158] Yang, B. A voice onset time comparison of English and Korean stop consonants. 20:41–59, 1993.

[159] Yanushevskaya, I., Gobl, C., Kane, J., and Ní Chasaide, A. An exploration of voice source correlates of focus. In *Proceedings of Eleventh Annual Conference of the International Speech Communication Association*, 2010.

[160] Yoon, K.-C. The role of prosody in dialect synthesis and authentication. *Journal of the Korean society of speech sciences*, 1(1):25–31, 2009.

[161] Ze, H., Senior, A., and Schuster, M. Statistical parametric speech synthesis using deep neural networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 7962–7966. IEEE, 2013.

[162] Ze, H., Senior, A., and Schuster, M. Statistical parametric speech synthesis using deep neural networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 7962–7966. IEEE, 2013.

[163] Zen, H., Nose, T., Yamagishi, J., Sako, S., Masuko, T., Black, A. W., and Tokuda, K. The HMM-based speech synthesis system (HTS) version 2.0. In *SSW*, pages 294–299. Citeseer, 2007.

[164] Zen, H., Tokuda, K., and Black, A. W. Statistical parametric speech synthesis. *Speech Communication*, 51(11):1039–1064, 2009.

[165] Zetterholm, E. Same speaker–different voices. A study of one impersonator and some of his different imitations. In *Proceedings of the 11th Australian Int.Conf. on Speech Science & Technology*, pages 70–75, 2006.

[166] Zheng, D. C. Accent Conversion via Formant-based Spectral Mapping and Pitch Contour Modification. *M.Phil. Dissertation, Dept.of Engineering, Univ.of Wolverhampton*, 2011.